

A council ripe for abolition

The British research council with responsibility for the natural environment could with profit be dismembered into smaller pieces.

THOSE who read the account of the collapse of the British Geological Survey's reason for existence on p.499 should prudently also refer to last week's complaint of maladministration at the Science and Engineering Research Council (*Nature* 4 October, p.397) and to the opinion expressed two months ago (*Nature* 26 July, p.261) that the British research enterprise is either bankrupt or about to become so. The circumstance in which an organization employing 1,000 professional geologists should be compelled, at least during the present financial year, to reduce to one-third its geological surveying is a scandal by any standards. If resources had been wasted on such a scale by a joint-stock company, and if they had been measurable in money terms, the people responsible would have been sacked. Because, in this case, the waste can be assessed only intangibly, by the frustration and disappointment of those concerned, by the lack of what they might have done and by the contagious effects of their sunken morale on other professional people, there is every likelihood that under the British system, the villains responsible will still be mentioned in future lists of public honours.

The origins of the scandal are clear. The Geological Survey of Britain began work in 1835, a short while after the Reform Bill of 1832 and a full three decades before *Nature* began appearing every week. For most of its history, the survey was administratively untouched, but in the early 1960s, what with the rhetoric of Mr Aneurin Bevan (since dead) on the need for a thorough exploration of what lies beneath Britain, and the fashionable pre-occupation with tidiness, it was inevitable that a government committee (under the late Sir Graham Sutton) should advocate setting up a research council concerned with the natural environment. The original idea was that the council should bring together the old geological survey, the Meteorological Office and various bits and pieces so as to forge an entirely novel instrument for multi-interdisciplinary investigation of the natural world, straddling ecology, the oceans and the atmosphere (not to mention the pollution thereof) as well as the solid Earth.

The Natural Environment Research Council (NERC) was already in a sense a failure when it came into being in 1967. The Meteorological Office had used its influence with the defence departments to opt out, with the result that the geological survey was the largest operational function with which NERC was involved. For many years, the new council did its best to combat the then expansionist tendencies of the geological survey (renamed for a time the Institute of Geological Sciences). Since the British Government's enthusiastic but uncomprehending adoption of the Rothschild recipe for the organization of civil science in 1973, the council has steadily won that battle.

Starting with the Rothschild precedent that government departments should pay for research they find worthwhile, NERC has sent the British Geological Survey (BGS) out into the market place, bidding for whatever odd jobs oil companies and minerals companies have wanted done. Nobody has ever examined the propriety of the accounting principles on which BGS has been able to win this business while still enjoying overhead support from NERC. The conflicts of interest that must arise when an agency whose prime function is to compile a public data bank (a geological survey) is also compelled to do jobs for private companies in return for cash have not been considered seriously. And NERC seems never to have asked the broader

question whether a country such as Britain needs a geological survey (the answer must be yes) and, if so, how it should be managed.

For those concerned, there are inevitably extenuating circumstances. NERC, from the outset fearful that one of its operational tails (the geological survey) would wag the dog (itself), and distract it from the general support of the Earth sciences, has consistently and properly been seeking to quarry more funds from its budget for the support of academic science. The survey, on the other hand, has fought hard and often arrogantly in defence of its unspecific goal of mapping the subsurface of the United Kingdom in a manner and to a timetable of which it is the principal arbiter. The trouble, for NERC, is that much else of its business is organized into independently operational units, but without clear objectives of what the operations are intended to achieve by when. There is, for example, a British Antarctic Survey (with a post-Falklands doubled budget), an Institute of Oceanography lacking formal connections with academic oceanographic research and an Institute of Terrestrial Ecology whose work (mostly good) is more easily described than its function understood. NERC has been beastly about the annual budget to all these organizations, no doubt winning other false economies like those effected at BGS. But NERC's failure to bring about a radically different pattern after nearly twenty years, exacerbated by the government's failure to recruit a full-term successor to Sir Hermann Bondi as NERC's chairman, points to a simpler solution.

Why not simply disband the Natural Environment Research Council, finding more suitable homes for the laboratories in its charge? That task should not be nearly as formidable as it seems. The geological survey is necessary and could be continued securely if it were the principal responsibility of a ministry, probably the Department of the Environment. The Institute of Oceanography should be integrated with academic research, while the ecology institute (three laboratories) should become the research arm of the Nature Conservancy Council (which, once, some parts of it were). It is not altogether a joke to argue that the British Antarctic Survey might go to the Foreign and Commonwealth Office.

Reorganizations such as these are no doubt more easily carried out with scissors and paste than on the ground but, for NERC and its masters, there seems no alternative to making the attempt. There is, however, a serious difficulty — what should be done with NERC's small but slowly-growing support for academic science? The obvious solution is to transfer this important rump to the Science and Engineering Council (SERC). The obvious snag is that SERC is not immune from precisely the kinds of troubles that afflict NERC, and is in any case already so huge that it dominates the whole British system for the support of civil science. Transferring responsibilities elsewhere (biology to the Medical Research Council, engineering to a new council as British engineers are forever asking) would not suffice. Transferring SERC's central facilities intended for university use to some trust that would use them on a European footing would be the ideal, but would not easily be traded off against the right of access to comparable facilities elsewhere. But while both NERC and SERC are in need of radical change (dismemberment and reform for the sake of flexibility respectively) the present is as good an opportunity as any to attempt this task. □



Australia's problem

Mr Bob Hawke's government must be positive about its support for research.

AUSTRALIA'S Labor Government has chosen the oddest way in which to prepare for the impending election, most probably in December. During two years of office, Mr Bob Hawke and his colleagues have been more or less committed to the view that Australia needs to make a special effort in industrial innovation if it is to be successfully dragged into the modern world. The prophet of change has been Mr Barry Jones, minister of science and technology in Mr Hawke's government, whose book *Sleepers, Wake!* has indeed stirred many Australians from their contentment. So why is it that in Mr Hawke's pre-election budget, published just over a month ago (see *Nature* 4 October, p.405) Mr Jones's department has fared worse than any other? The cash increase does not match inflation, so that the resources available will actually be reduced. So what will happen now, both to Mr Jones and to his policy?

Australia, with a population half that of California, is economically the victim of its rich natural resources. For decades, Australians have been arguing the toss about the efforts they should make to add value to the primary products, in agriculture and minerals, which as exports have bought modest prosperity for what used to be the most successful practitioners of Western ways in the South and West Pacific. The standard political response has been to restrict overseas investment in Australian primary production, and to protect Australian manufacturing industry (which has never seriously caught on) from outside competition. Mr Jones's tract was meant as a plea for more adventurous initiatives. Why not recognize that the world is changing, and becoming more dependent on the self-conscious application of technology? Why not put the country's sophisticated technical community to better use, both by means of links with novel industries but even in traditional fields such as agriculture (as by the self-conscious encouragement of biotechnology)? And why not (because there is no choice) confess that Australia must make its way in the world in partnership and competition with the other economies of the Pacific basin, which include Japan as well as California?

While some of Mr Jones's tub-thumping has occasionally been naive, it is a shame that he has not been listened to more attentively, both by his colleagues in the government and by the technical community. The trouble with Australia, whatever the political labels its governments may be wearing, is that it is deeply conservative. (Changing governments is permissible, but changing the ways in which people earn their livings is something else.) This cast of mind is reflected in the pattern of Australia's research effort, which is substantial (for a primary producer). The Commonwealth Scientific and Industrial Research Organization (CSIRO) is eminently suited to the improvement of traditional agriculture and the more effective exploration for minerals (with substantial side-interests in such fields as solar energy), but is so heavily institutionalized that it cannot be the radical force that radical Mr Jones has been looking for. During his spell in office, Mr Jones has done his best to loosen the bonds of Australian self-satisfaction, and there have been some signs of success. But these do not amount to the radical change for which Mr Jones has been working.

What will happen next is anybody's guess. Mr Hawke may not be re-elected, in which case Mr Jones's ambition will for a time be forgotten, but that is not what the pollsters say. In the hope of salvaging something from the disappointments on the technical front during the past two years, it is reported that Mr Hawke is planning a great administrative initiative — some kind of super-ministry that will unite the Australian Government's responsibilities for research and for the support of industry. That will be a strictly cosmetic device, and one that must be meaningless on the eve of an election. It would be more profitable for Australia and for the present government, even at this late hour, if Mr Hawke and Mr Jones could work out a less negative plan than that

implicit in the budget, whose objective seems to be to teach the technical community a lesson the hard way — cutting back on resources through existing channels with the unrealistic objective of forcing people into partnership with an industrial sector notoriously unwilling to spend resources on research. The government is right not to try to back winners in a horse-race not yet devised, but wrong to think it can conjure industrial support of research simply by cutting back on what it spends itself. Far better, in reality, to embark on selective support for imaginative work at the universities and CSIRO institutes, so providing a showcase of what may already be available. A sum of money not much greater than the \$A30 million set aside for the defence of the America's Cup would go a long way to help. □

Conservation cheers

The tide is running in favour of British conservation.

THE British conservation movement, a little to everybody's surprise (its own included) is in good spirits. So much is clear from the tone of the tenth annual report of the Nature Conservancy Council (published on Tuesday), which gives an account of a frustrating year without undue wringing of the hands. Evidently the council's confidence has been reinforced by what others active in the field have been saying. The House of Lords committee which a few months ago pointed to the conflict of interest between agriculture and conservation, and asked that it should be redressed, has been good for people's morale. The Countryside Commission, like the Nature Conservancy Council a statutory body with formal functions defined by the Countryside Act (1981), has also been strong and sensible about the importance of the appearance of landscapes of all kinds. Already the council appears to have forgotten last year's rows, when the then chairman found that the government's refusal to renew his appointment did not excuse him from his duty to support an agriculture minister in a confrontation with farmers angry that ploughing a marsh in Somerset might be forbidden.

The most astounding development, however, is the noises being made by the National Farmers' Union, hitherto hot-gospellers of hedgerow destruction (with financial help from the British Government), contour violation by access roads (again grant-aided in the hill regions that are most easily damaged) and the replacement of traditional buildings by metallic barns designed to be seen from as far away as possible. The union's officials are now talking about the need to abate the rush for farm production by some consideration for conservation. Even if, as may be possible, the farmers have not repented of their ways but have merely calculated that it will be the worse for them if they alienate all countrygoers, the development has been a heartening surprise. The danger now is that people will be lulled into complacency.

One obvious trap stands out from the Nature Conservancy Council's report, which repeats the now familiar tale of its difficulties in administering the section of the Countryside Act allowing the designation of "Sites of Special Scientific Interest". When properly designated, sites cannot be altered radically by their owners (who may nevertheless have to be compensated for income forgone, or even bought out). The practical difficulty is that the council, with its small staff, cannot cover the ground as quickly as it wishes or as prudence requires. In the three years since the act became law, a third of 4,000 sites have been dealt with, but those remaining include some of the hardest cases. The council is now complaining of loopholes in the law, but it should seriously consider whether it should not restrain its ambitions for notifying so many sites. It has other powers (to protect species or create nature reserves) which might often serve the same purpose. If the tide of opinion is flowing in the right direction, there is less need to hurry. And it is asking for trouble to urge, as in this week's report, that there should be nationally agreed percentages of consumption of temperate foodstuffs to serve as permanent targets for the farming community. □

Genetic manipulation

Rifkin strikes against gene transfer experiments

Washington

ANTI-genetic engineering activist Jeremy Rifkin once again thrust recombinant DNA research in the United States under the spotlight of national attention last week. Rifkin's lobbying organization, the Foundation on Economic Trends, together with the Humane Society of the United States, has filed a suit in the district court here to halt gene-transfer experiments being carried out in livestock by the US Department of Agriculture (USDA). The plaintiffs allege that the experiments are a violation of the ethical canons of civilization and represent a new form of cruelty to animals.

The experiments entail injecting fusion genes that include the DNA structural sequence of the human growth hormone

genetically engineered livestock, with small farmers being forced out of the market and higher prices being passed on to the consumer. The Humane Society of the United States argues that physiological imbalances in genetically engineered animals would cause suffering.

Brinster, evidently feeling besieged, was trying to avoid the press last week, but wants the public to understand the possible benefits from the research. Apart from the prospect of better and cheaper meat and other animal products, gene transfer experiments hold great potential for cancer research, especially with Leder's recent demonstration that transgenic animals can inherit a propensity for particular cancers.

The human growth hormone sequence was used simply because it is well studied and is easier to detect in mice than the rat sequence originally used (see *Nature* 300, 611; 1982). Despite the difficulties of working with opaque eggs, the livestock system is working well. Brinster declines, however, to discuss preliminary findings

until definitive results are ready for publication. So far "several hundred" eggs have been injected with the growth hormone fusion gene; if successful, the experiments will allow further work on the processing of gene products, because livestock animals provide larger tissue samples than mice.

USDA officials concede that Rifkin's challenge will probably be pursued by the court, but they have not yet studied the litigation in detail. Whichever side loses in the first round of the action is certain to appeal, and the case could drag on for a long time.

The three causes of action cited by the plaintiffs refer to the creation of federal common law nuisance and violations of the National Environmental Policy Act and the Administrative Procedure Act. In essence, the claim is that USDA's animal breeding programme is at odds with policies approved by Congress in other areas, and that USDA has not adequately considered other strategies or the impacts of its programmes. Rifkin is also trying to introduce an amendment to the guidelines used by the Recombinant DNA Advisory Committee of the National Institutes of Health, which would restrict research in the area and which due to be discussed later this month. In the meantime, the research continues.

Tim Beardsley



into the fertilized eggs of sheep and pigs. When similar experiments are done with mice, some of the resulting animals express the human gene, and so grow larger and faster than normal mice (see *Science* 222, 809; 1983). The new work is a collaborative effort between Dr Ralph Brinster of the University of Pennsylvania Veterinary School and a group under Dr Vernon Pursel at USDA's agricultural research centre at Beltsville, Maryland.

While the human origin of the growth hormone sequence doubtless added to the public's fascination, Rifkin makes clear that his objects apply to transferring any foreign genetic material into the germ line of a species. It would be morally objectionable to transfer animal genes to humans, Rifkin says, and it is similarly objectionable to transfer human genes to animals. Furthermore, the biological stability of the species would be threatened if farmers switched in large numbers to "super animals" containing genes from other species. Rifkin believes there would be undesirable economic and environmental consequences of the general use of

Dutch universities

Small print conceals more cuts

THE Dutch universities will feel the pinch again next year, despite ministry of education figures that make it appear that university research spending has increased substantially. The ministry has achieved this sleight of hand by recalculating the time a university teacher spends on research.

In the 1970s, the ministry assumed that natural scientists in the Netherlands spent 36 per cent of their time on research — and their salaries were debited from the research budget in proportion. In 1985, however, the assumption will be very different; research time is estimated to be 44 per cent of duties on average, varying from 18 per cent in academic hospitals to 61 per cent in mathematics.

The result in the nominal university science budget is that what appears to be an increase, from Dfl 1,452 million (£341 million) in 1984 to Dfl 1,796 million (£418 million) in 1985, is actually a decrease in real terms. If the budget is calculated using the 36 per cent research time figure, and compared at constant prices with the 1980 figures, it appears that the 1983 budget was just 93 per cent of the 1980 figure, the 1984 budget 92 per cent and next year down to 91 per cent. Of course the new calculations (which put the 1985 university research figure at 114 per cent of that in 1980, at constant prices) do not make more money available to research, and by raising the apparent budget make it more vulnerable to

the threat of cuts at cabinet level. However, the re-calculation was merely for administrative convenience, and not a policy matter, a government spokesman said last week. The new research-time figures are based on careful surveys of what Dutch researchers actually do, and are not pulled out of a hat. Interestingly, though, they conflict with lower fractions generally assumed in other countries. In Britain, for instance, the University Grants Committee estimates research time to be just over 30 per cent of duties.

Robert Walgate

● Total Dutch spending on research and development will increase slightly next year to Dfl 8,132 million (£1,890 million), according to recent government predictions. Government spending will be Dfl 3,882 million (£903 million), a slightly lower fraction of the Gross National Product (GNP) (0.97 per cent) than last year (0.98 per cent), but still larger than in the years 1980–83 (0.90–0.93 per cent). Industry will spend Dfl 3,800 million (£884 million), the government estimates, and another Dfl 450 million (£105 million) will go to international projects.

Dutch research spending is therefore below the European Economic Community average as a fraction of GNP, but only 3 per cent of the Dutch research and development budget is spent on defence. This compares with 33 per cent in France.

Casper Schuurin

French medical institutes

Twelve-year rule begins to bite

THE French medical research council, INSERM, is approaching its twentieth anniversary — and a time of testing. A government attempt to stop ageing laboratory directors sleeping at their posts, known as the “12-year rule”, takes effect next year — when all who have been directors for 12 years or more must resign. Thus 60 out of 250 INSERM directors must go or reapply, which may explain why protests over the “bureaucratic” rule have been more bitter at INSERM than at any other French research council affected.

At the Centre National de la Recherche Scientifique (CNRS), for example, there has been little resistance — perhaps because, despite its size and age, CNRS has only 50 directors affected by the rule. This is attributable to physicists who often change directorships voluntarily before the 12 years are up, and to biologists who have been applying a policy of group concentration and reorganization, thus anticipating directorship problems. Altogether, CNRS seems naturally to have shown more of the flexibility that the 12-year rule is intended to encourage.

Flexibility will come, even at INSERM. Director-general Philippe Lazar is privately delighted that the “majority” of the 60 affected directors have not yet re-applied (with revitalized research programmes), as they are entitled to do by mid-November. This should give him some fun next year, when he will be able to reallocate this research potential according to new priorities now being defined by his advisory science council.

At INSERM, there were two kinds of resistance to the rule: from conservatives, whom — broadly speaking — the rule was designed to shift; and from a group of first-rate directors who are likely to remain directors after reapplication, but who deplore the paperwork and time involved. But among the latter, many have found the exercise of redefining 12-year research programmes “refreshing”, Lazar claims.

Lazar, a medical statistician, has set out to refresh INSERM since he took over in 1982. “The image of the ivory tower has gone”, he says. The 12-year rule is important to him because it makes it possible for directorships to change hands without the slur of explicit negative judgement on the incumbents.

Another major change at INSERM during Lazar’s term has been a shift towards health service research, one outcome of an “honest review of strengths and weaknesses” instituted by Lazar and completed this summer. Some of the changes expected by next spring include:

- The creation of a new scientific committee (there are at present eight) in “the human and social sciences applied to health” with the objective of establishing new research fields and units.

- A general reorientation of priorities for new units established by INSERM, and a similar reorientation of INSERM external grants.

- An improved recruitment system for new scientists — with 60 new posts next year.

Several problems remain unsolved, however. There should be more foreign experts in review committees, Lazar admits, but “we shouldn’t expect that foreign experts will do miracles”. In INSERM itself, there are “silent experts” who refuse to judge their colleagues, but that is changing. And while INSERM (like the whole of France) is hamstrung by the total lack of postdoctoral fellowships, plans are under way at the research ministry to introduce a fellowship system.

Otherwise, the resolution of one old problem may mean less cash for French

biologists. In its early days, INSERM competed openly with CNRS for the best biologists, to the benefit of many. But under a new agreement between the councils, that boon will end. Laboratories which have been jointly funded for no good policy reason will now have to adhere either to CNRS or INSERM, probably with a consequent loss of funds after a four-year transition period.

But to replace the past *ad hoc* joint support, there will now be three clear structures for collaboration: a few (10–20) existing common laboratories where the basic research goals of CNRS clearly overlap with the medical goals of INSERM; new “concerted development units” where INSERM and CNRS wish actively to set up joint action and “allied units” where, for example, CNRS staff may be working in an INSERM laboratory, or vice versa.

Thus clarity is entering the INSERM–CNRS relationship for the first time in twenty years, and French biomedical belts may tighten as a result. **Robert Walgate**

Soviet geology

Deep boreholes yield data-harvest

ON 7 September 1984, drilling operations began on a projected 15-km borehole through the Ukrainian crystalline shield from a site at Novoivanovka, near Krivoi Rog. This drilling project, involving more than 150 research institutes and “production collectives”, marks a new stage in the Soviet deep borehole programme — the sinking of deep shafts specifically related to the search for metals.

Soviet interest in deep boreholes dates back to the beginning of the 1960s when the United States was planning to drill a “mohole” into the oceanic crust down to the Mohorovicic discontinuity between crust and mantle. Soviet plans have concentrated on drilling of several deep boreholes (12,000–15,000 km) at carefully chosen sites — the Kola peninsula, the Kura depression in Azerbaijan, the Urals, the Caspian depression and one of the Kuriles. At the outset, however, Soviet drilling technology was insufficiently advanced to reach the proposed depths. Oil drillings averaged less than 2,000 metres in depth, and the few shafts that passed the 4,000-metre mark were considered exceptional.

Not until May 1970 could work be started on the first deep hole, at a site 8 km north of Zapolyarny on the Kola peninsula, while the second, at Saatly in Azerbaijan, was started only in June 1977. Special drilling rigs developed by the Uralmash plant differ from US deep-drilling equipment in that the string remains virtually stationary while the bit, a miniature turbodrill, rotates. The special-purpose instruments, capable of withstanding high pressures and temperatures, include a miniaturized motor, developed by students at the Kaunas Polytechnic

Institute, which can drill and analyse rock samples at depths of over 11,000 metres.

Already the programme has yielded interesting results. Samples from the Kola borehole have revealed in the Baltic shield aqueous solutions of carbon dioxide and helium at depths of 6,000 metres, the expected granite-basalt boundary at approximately 7,000 metres was found not to exist and the angle of inclination of beds which form outcrops of 45° was found to remain the same at increasing depths, contrary to the expectation that the inclination would increase. Fossilized microorganisms were found in strata believed to be 2,000 million years old. Fossils from the Saatly hole (shells and small corals) have confirmed that there had once been a shallow sea in that area.

In the Kola borehole the temperature at the 12,000-metre mark (on 28 December, 1983) was above 200°C, instead of the expected 100 to 120°. In the Saatly hole, the temperature at 7,500-metres was 140°C, less than expected, suggesting that there is no deep magmatic activity in the Kola depression.

The practical implications of the data have proved more valuable than expected. In the Kola borehole, traces of copper, lead, zinc, cobalt and nickel have been found far deeper than predicted, so the programme has been extended to include a further 20 holes in ore and hydrocarbon bearing localities. Boreholes in oil/gas-bearing regions will include sites in the Tyumen region, the Kuban, the north Caucasus, the Caspian depression and the Timan-Pechora region. The four metal-ore sites will be in the Urals, near Noril’ski in Uzbekistan and at Novoivanovka.

Vera Rich

Polish universities

Autonomy beyond reach

POLISH universities opened last week in somewhat gloomy circumstances. The Party daily, *Trybuna Ludu*, greeted the new term by observing that "for several years, the intellectual activity of the academic milieu has been declining" and has failed to produce "fresh achievements significant in world standards".

At the same time, the paper gave a strong hint that the autonomy of the universities was, essentially, part of an implicit bargain by which the university authorities, in their turn, assumed "co-responsibility" to prevent university self-governance being used as a cloak for "harmful" actions.

Neither of these criticisms is unexpected. The declining standards of Polish science have been repeatedly discussed, explained and excused in the Polish press during the past year, and may fairly be summed up as the result of the hierarchical planning of research during the Gierk regime (when every research project had, officially, to relate to a carefully graded ladder of "problems" of the national economy). But, now, all foreign currency for the purchase of equipment and journals has suddenly been cut off. The situation is so bad, a working party of the Polish Academy of Sciences warned last month, that without an immediate injection of hard currency, Polish science could come to a complete halt.

University autonomy has also been a contentious issue during the past year. The chief clash was over the election of university rectors, when three rectors-elect were vetoed by the Minister for Science, Higher Education and Technology, Dr Benon Miskiewicz; when the Warsaw University electoral college refused to accept his decision, its functions were suspended for six months and the minister let it be understood that he would appoint a rector of his own choosing. Yet when the university opened last week, the outgoing rector, Dr Kazimierz Dobrowolski, presided over the beginning-of-term ceremonies and, in his inaugural address, expressed the hope that the university would soon have a new rector, chosen by the electoral college.

It is now thought that Dr Dobrowolski came to an agreement with the minister during the vacation, by which he would continue in office until 15 February 1985, and that the electoral college would elect his successor when its term of suspension expires in February. But none of this has appeared in the Polish press, and there is no certainty that, even if there is another election, the minister will not again intervene if the choice is too "daring".

The issues of autonomy and of research standards are, in fact, related, since one of the major demands for academic autonomy formulated in the Solidarity period was the right of universities to determine their own research programmes, inde-

pendently of state priorities. But even the autonomy granted by the 1982 Higher Education Act (considerably less far-reaching than was originally hoped for) is not always easy to implement. In a recent interview in *Polityka* the rector of the Jagiellonian University of Krakow, Dr Józef Gierowski, noted that university self-government cannot be invoked to deal with the bottom-heavy structure of academic teaching staff, where too few assistants and junior lecturers are prepared to qualify themselves for professional posts.

The university, he explained, "being as it were a continuation of the function of the welfare state", is obliged to employ "scientific workers", for whom the PhD degree represents the "peak of their capacity or

ambition". (Senior academic staff in Poland must hold the higher degree of *doctor habilitatus*). The only success for "autonomy" in dealing with such people is the acceptance of the principle that they must resign if they do not get their doctorate within eight years.

Polityka, which, although less liberal than before Martial Law, still enjoys considerable prestige, has during the past year given considerable space to problems of science and education. Recently it has been particularly concerned with the training of Poland's future scientific workforce, covering such issues as the falling interest among university entrants in the technical engineering sciences, the challenge of the "third wave", the relevance of "new maths" in primary schools, and the practice, recently introduced in some schools, of a longer school-day with Saturdays free.

Vera Rich

US-Japanese education

One man's meat, another's poison

Tokyo

JAPAN and the United States have agreed to set up a joint 30-member council to compare their educational systems from May of next year. The move comes not before time, for both nations have been worrying about the weaknesses of their respective systems (helped along by reports of emerging IQ differences) and envying the strengths — real and imagined — of the other's.

At the Washington meeting between Japan's Education Secretary Yoshiro Mori and US Secretary of Education Terrel H. Bell that gave approval to the council, Bell, reflecting growing US desires for a return to a more disciplined educational system, lavished praise upon Japan's tough *juku* (private cramming school) system, saying that the United States "must have *juku* because the Japanese are so productive".

A horrified Mori then had quickly to explain that the Japanese Government, reflecting growing desires for a more liberal educational system, regards private cramming schools, at which some students spend four hours a night, as an evil produced by the terrible competition for entry to leading universities and high schools, and hopes to do away with them. Later, Bell, well known as a supporter of increased homework in US schools, admitted that he had not actually known what *juku* were but what he really wanted was a longer school day "which is a supervised, studying, tutoring, *juku*-type thing" (*sic*).

The US-Japan council will have its first meeting in Tokyo and will have a special panel looking at problems of science and mathematics education — of particular concern to the United States. Experts from the United Kingdom, France and Germany are also to be invited.

Shortly before Mori's meeting with Bell,

Japan's Prime Minister, Yasuhiro Nakasone, finally succeeded, after six months of Diet wrangling (see *Nature* 308, 678 and 309, 739; 1984), in inaugurating the special council which is to spend three years devising reforms of Japan's education system.

Despite vociferous protests from the socialist and communist parties, the council has been handpicked by Nakasone and is answerable directly to him. Only two school-teachers are among the council's 25 members (average age 60) and there are no representatives from the nation's biggest teachers' organization, the Japan Teachers Union, which virtually ensures that conflict will continue. The post of president of the council has, however, gone to a well known liberal, Professor Michio Okamoto, until recently president of Kyoto University. Nakasone has asked the council to take speedy action on the biggest problem — the university entrance examination — and to seek "diversity" in the education system.

Exactly what diversity means is likely to cause considerable debate. In the present straight "one-track" educational system, completely unstreamed groups study the same subjects and try the same examinations, with only total scores deciding how a distinguished university faculty can be entered (and thus often what subject is studied). The liberals would replace it with one in which students have freedom to follow their own interests and receive an education tailored to their own particular abilities, but what the left wing fears is that "diversity" will simply mean increasing the scale of divisions between top quality (and expensive) private schools and the poorer state schools: an effective return to a pre-war system where elites were selected for special training at an early age.

Alun Anderson

Japanese universities

Broadcasting for degrees

Tokyo

JAPAN'S first broadcasting university — the University of the Air (Hoso Daigaku) — is poised to go into action: its buildings and facilities are just being completed in Chiba City, near Tokyo, after an initial expenditure of 10,000 million yen (US\$420 million). Application forms for the enrolment of its first 10,000 students will be available in a few weeks. Test broadcasts will start at the same time, beginning with programmes explaining what the university has to offer and how to apply for entry.

Once in full operation in April next year, the University of the Air will become the first broadcasting university in the world to have its own specially dedicated television and radio channels, each carrying lectures from 6 a.m. to midnight seven days a week. Adding to this generous helping of learning the 18 hours of school, documentary and language programmes broadcast each day by the national television corporation (NHK)'s Channel 3, there will be a little over 250 hours of television tuition available each week — enough, perhaps to satisfy even the Japanese thirst for education. The university's programmes will, however, be available only to those in the area around Tokyo, for plans to broadcast directly by satellite fell through and conventional transmitters are being used; the rest of the country will have to wait another five years for the service.

Support for the establishment of a broadcasting university resulted partly from the calls for fundamental educational reform after the 1960s student unrest. At Tokyo University, the centre of the trouble, teaching was totally disrupted, the entrance examinations had to be cancelled (losing a year's student intake) and the general public were treated to television coverage of the nation's young elite flinging "Molotov cocktails" from the university tower at riot police below.

With the ending of the troubles, no fundamental reform of university education was carried out, but some new educational projects did get under way: the new "American" style university at Tsukuba (see *Nature* 305, 371; 1983); a pair of colleges to which school-teachers can return for two years of "refresher" graduate-level training at the government's expense (provided they can pass the tough entrance examination) and, now, the University of the Air.

Despite the ¥10,000 million government expenditure on the university and the fact that the Ministry of Education, Culture and Science (MESC) will pay its operating budget, the government will have no direct legal control over the university, for Japanese law guarantees independence of the mass media. Instead, the university is run by a foundation, and its staff, unlike those of a normal university, will not be

civil servants.

And although the university buildings are right alongside a new National Centre for the Development of Broadcasting Education, a part of MESC from which the university is renting its broadcasting studios, the university building has been kept just a few metres away to demonstrate its separateness. Only a corridor connects the two. But this arrangement has not fully satisfied the socialist opposition parties, and courses such as sociology and education which can encourage critical thinking about government policy are certain to be carefully watched.

The overall teaching system has much in common with other broadcasting universities. Set texts accompany lectures which are broadcast several times a week at different times of day, tests and reports are sent by mail and face-to-face teaching and end of term examinations given at six regional study centres. Success in the examinations gains credits towards the 124-credit total, and minimum of four years' study, needed to gain a degree. A typical two-credit course requires the understanding of a 120-page book and 15 accompanying 45-minute lectures: it is expected that students will average five broadcast lectures a week plus three hours of lectures at study centres.

Three main disciplines are to be offered: science in everyday life, industrial and social studies, and humanities and natural sciences, but within this framework there is to be considerable choice of specialized subjects after fundamental studies have been completed. Of the 10,000 students to be enrolled to begin with, 4,000 are expected to take a degree course (high-school graduation is the only requirement) and the remainder to take a few subjects for a short period. A full degree course will not be free, although at a total cost of ¥750,000 (\$3,000) it will be cheap compared with attendance at a university.

Around 50 professors have been chosen to run the university and will be backed up by an equal number of associate and assistant professors. Not having civil servant status, however, the junior staff will be employed on renewable five-year contracts rather than given tenure. Some 100 guest lecturers will also be brought in from other universities on contract to make particular television and radio programmes.

Efforts will be made to ensure that the University of the Air degree is a real degree and not a second-class qualification. To this end, the university is now trying to arrange for credits to be transferable to and from other universities. If this proves successful, given the notorious lack of interchange of students between conventional universities, it will be a remarkable achievement in its own right.

Alun Anderson

US basic research

Cheer and gloom at war

Washington

SHARPLY contrasting views on US university research were aired at a meeting held last week by the American Council on Education. Dr Edward David Jr, president of Exxon Research and Engineering and a former presidential science adviser, told the conference, held at the University of Georgia at Athens, that the growing links between university research and industry were "a promising picture for the future". But Professor William McElroy, professor of biology at the University of California at San Diego, said universities were facing "almost a disaster" in research support.

In his opening speech, David argued that the public's desire to create jobs based on new science, together with the power of new instruments, will lead to more cooperation between universities and private research institutions: industry has been the fastest growing source of support for university research since 1970. But David was concerned that manpower competition from large federal projects in defence and space might crowd out the civilian sector.

McElroy said bluntly that such optimistic views "do not stand up against the facts". Despite constant-dollar increases in federal university research support since 1972, the escalating cost of doing research has meant that the same money now buys less. The amount spent per grant has decreased substantially, and it is becoming harder to find money for research projects: five years ago, a merit rank of 3 on a scale from 1 to 5 would have been sufficient to ensure support from the National Institutes of Health, but now a rank of 1.5 would be insufficient. And McElroy commented that industrial support for university research is concentrated in relatively few universities.

McElroy's remarks echo comments made recently by Donald Kennedy, president of Stanford University, in a letter to Dr George Keyworth, director of the White House Office of Science and Technology Policy. Kennedy wrote that undercapitalization of teaching and research laboratories is causing problems in retaining faculty, while the need for renovations of research facilities had "reached crisis proportions".

University research is the subject of a major study nearing completion in the White House science office, chaired by David Packard, president of Hewlett-Packard. The study is rumoured to back the case for more generous support of basic science by federal agencies. And the Department of Defense is already preparing to step up its programme of unclassified research in universities.

Tim Beardsley

British geology

Strategic survey grinding to halt

GEOLOGICAL surveying of Britain by the British Geological Survey (BGS) has almost ground to a halt this year. Because of over-dependence on short-term contracts and the cost of its recent centralization at a site in Keyworth, Nottinghamshire, strategic surveys of land are running at only 30 per cent of last year's activity.

At the same time, BGS seems to have lost its old friends. Private industry is complaining about its function and performance, BGS has no long-term commitment from government departments that use its services and, unlike most national surveys, it has no direct responsibility to a government minister for strategic surveys and information management. With its parent body, the Natural Environment Research Council (NERC), itself under severe financial pressure, BGS has an uncertain future.

The decline of the land survey activities comes at a bad time. Since 1835, BGS and its predecessors have been preparing geological maps of the United Kingdom at 1:50,000 scale (1 inch to 1 mile), with more detailed mapping following later. Although most of the United Kingdom has been mapped at 1:50,000, at least 25 per cent of the maps are obsolete, according to BGS. Moreover, at the more detailed 1:10,000 scale, large areas of the country remain unmapped, partly because of concentration on regions important for economic or development reasons.

After the survey was attached to the newly-formed NERC, in 1965, it accumulated a wide range of new responsibilities: overseas activities, geochemistry, geophysics, seismology, geomagnetism, mineral and water resources. When, in 1973, the Rothschild "customer-contractor" principle was accepted by the British Government, a portion of BGS funds were transferred to government departments. By 1979, the land survey activities were financed and controlled by a consortium — 60 per cent NERC, 30 per cent Department of Environment, 5 per cent Department of Trade and Industry, 5 per cent Department of Energy. Meanwhile, other parts of the survey were so heavily engaged on commissioned work that three quarters of the survey's funds came from the customer departments.

Although cutbacks by government departments began in 1980, the first major blow to the land survey was struck in 1982 when the Department of the Environment withdrew from the consortium as well as from other activities, replacing that commitment by a series of short-term (six months) contracts whose allocations were not made until after the start of the financial year. The Department of Energy, meanwhile, transferred its consortium support to deep geological research related to onshore hydrocarbons assessment.

These developments left the land survey almost entirely dependent on NERC. The latest blow, earlier this year, was a cut by NERC in BGS science funds of £1 million (about 10 per cent) made necessary because of NERC's severe financial shortage. Because short-term BGS commissioned contracts draw on this science budget, and because the recent move to Keyworth has entailed relocation expenses, the result has been a 70 per cent decrease in the funds available for scientific work. Now, routine activities such as geomagnetic monitoring

BGS budget £ million 1984-85 prices			
	1982 -83	1983 -84	1984 -85
NERC Science Vote	63.1	65.1	65.9
Commissioned income	30.8	26.4	23.7
BGS Science Vote	9.8	10.4	9.2
Commissioned	18.5	16.9	14.9

BGS Staff numbers:

1973: 757 1979: 1,156 1985 Target: 920

are threatened while the land survey programme has been curtailed to 30 per cent of last year's when, according to the director of BGS, Professor Malcolm Brown, this programme needed to be doubled.

There are those, inside and outside BGS, who conclude from this catalogue of misfortune that the Rothschild system was the biggest disaster to hit the survey since its foundation. There are others who say that BGS's problems are more fundamental.

Perhaps the strongest criticisms come from local government and private industry, which charge that the survey is simply not providing an adequate supply of basic geological information needed for planning purposes, for example. And the move to Keyworth will exacerbate the situation, the critics say, by rendering the available information less accessible. But BGS intends to maintain its London library until 1986 at least, and thereafter to maintain a London information centre with two BGS scientists and a duplicate supply of key geological maps and reports.

Many of the private companies are deeply irked that, as they maintain, BGS is in effect a state-subsidized competitor for short-term contracts which is too heavily engaged in such work to provide British industry with basic services. It is hard to document the charge, but staff at BGS concede that customers will tend to benefit from the large infrastructure available to support contracts placed with BGS.

Staff at BGS are all too aware of the backlog in the publication of maps and memoranda. One explanation is the diversion of resources and funds to other activities. It is estimated that one map at 1:50,000 scale now requires anything up to £1 million to produce from scratch. In fact, the survey has about 50 geological memoirs in preparation, some of which might be

published for as little as £20,000, but the money cannot be found.

Another contributory factor, put forward by a visiting group commissioned by NERC to assess BGS, is that while the collapse of the land survey consortium was much to blame, BGS administration of the preparation of reports and maps was inadequate — a charge also made by the previous visiting group a few years earlier. A long-term response to such criticisms, but one of BGS's top priorities, is the establishment of a national computerized database — a massive task, particularly now that strategic surveys use geophysical as well as geological techniques to prepare maps.

Complaints against BGS are offset by industry's acknowledgement of the value of services such as the provision of minerals intelligence and general commodity expertise. In 1981, minerals and quarrying (including oil) accounted for 18 per cent of UK production, and BGS supplies information ranging from distributions of potash horizons based on North Sea drilling to details of small barytes mines.

Ultimate — and, in many people's opinion, excessive — control of BGS rests with NERC. Academic geologists and those working for other national surveys are surprised, even outraged that the director of BGS has to refer to NERC headquarters each new appointment to his staff, for instance. Similarly, every decision relating to the purchase of computing equipment is referred to NERC's central division. Coincidentally, computing problems have been highlighted by a decision by the Department of Energy to cancel its contract with NERC for computing services associated with its continental shelf programme, and by a recent internal NERC report which recommended some decentralization of computing power.

NERC says its control of manpower, budget and the scientific programme of BGS stems from its accountability to Parliament. But NERC is aware of the backlog in the preparation of maps and memoranda, and has proposed that resources might be diverted within BGS for the work.

In short, there is general agreement that the function of BGS which is most in demand — the provision of basic information about the geology and natural resources of the United Kingdom — has, for a number of reasons, virtually fallen by the wayside. Some members of other national surveys suggest that BGS should be wrested from NERC and made responsible to a "minister for natural resources". As no such post exists in the United Kingdom at present, this radical proposal is unlikely to be realized, at least unless NERC itself is disbanded. Given the tight budgets within which BGS is operating, however, it is hard to see that a simple redistribution of its internal resources can have a significant impact on a central problem.

Philip Campbell & Peter Gambles

UK atomic energy

Putting research out to contract

THE UK Atomic Energy Authority (AEA) may soon be dead, but long live atomic energy in the United Kingdom. That seems to have been the mood at Britain's nuclear establishments last week, in the wake of the publication a week ago of a set of proposals for reorganization that will eventually saddle electricity consumers and other users with much of the cost of applied research in nuclear energy.

The new arrangements have been put to the Secretary of State for Energy, Mr Peter Walker, by an anonymous committee of officials from the department. The good news for AEA is that the committee has shrunk back from the notions that the authority should be dispensed with ("an authoritative body of expertise . . . will continue to be required") or transferred to the private sector (which is "not realistic").

Big changes are nevertheless intended. The committee recommends that a greater share of AEA's research should be paid for by the electricity industry, and that other research commissioned by the Department of Energy should be paid for on the "customer-contractor" basis.

The case for some continuing support for nuclear research is that the "government's policy that [Britain] should have a significant nuclear power programme for the foreseeable future" makes some kind of research programme necessary, while the "unique nature of nuclear power and public perceptions of it" will require continuing public support for programmes of research going beyond strictly commercial interests.

On the authority's specific research programmes, the official committee recommends that the electricity industry should shoulder a larger proportion of the cost of fast reactor design as research gives way to development, that work on thermal reactors should be paid for exclusively by the electricity generators, that there should be further discussion about sharing the cost of nuclear safety research, but that the nuclear waste programme should continue to be paid for centrally.

The committee says that the "underlying programme" of research has been of benefit to user organizations but that, on the principle that present customers are "proxies for future customers", they should now begin to contribute towards the cost of this research.

The next step in the unwinding of AEA's dependence on public funds is that the minister should decide whether to accept the proposals his officials have put to him, which will in turn depend on negotiations between the potential customers (which may be unwilling to pay for service hitherto free) and the potential contractor (which may also be an unwilling partner).

The Department of Energy is nevertheless hoping that the outstanding issues of

who pays what will be settled within a few months, which would have the advantage that much of the cost of the British nuclear research programme would disappear from the government's expenditure accounts before the beginning of the next financial year in April. Another specific recommendation of the committee, which has been generally accepted, is that AEA should in future be managed as a "trading fund", a public corporation subject to the accounting rules that apply to private com-

panies (and carrying the freedom to borrow from the banks). But this arrangement will require legislation, so the formalities are likely to be delayed.

In the authority, officials and staff members say they "have seen this coming for a long time". There are grounds for optimism that, in the short run, the authority (which already supports a third of the research at its largest establishment at Harwell through outside contracts) will be able to maintain the present level of activity. The dangers may come later, when customers for its research services may spend their contract money on quite different things. □

UNESCO

One friend for M'Bow

MR Amadou-Mahtar M'Bow, director-general of the United Nations Educational, Scientific and Cultural Organization (UNESCO), went last week to Strasbourg to defend his beleaguered organization's record before the Assembly of the Council of Europe. And while he had little success in convincing delegates that UNESCO's house is in order, the Council of Europe agreed that its secretariat should go ahead with plans to increase cooperation between the two organizations.

This move may stem from a fraternal feeling between UNESCO and the Council, both of which were set up after the Second World War as institutional symbols of peace and goodwill. Now UNESCO is in trouble, accused of mismanagement, and threatened with the withdrawal of the United States, the Council wishes to come to its aid. In fact, there have been regular contacts between the two organizations since 1952, when the first cooperation agreement was signed, and the plans in hand are merely to extend the contacts a little. The new agreement is to cover areas of common interest such as communications (satellites and copyright questions), cultural development and North-South cooperation. The vexed issue of "a new world information order" for news reporting is not among topics for discussion.

Even so, the mere presence of M'Bow at the Council's building in Strasbourg was too much for conservative group members of the assembly, who issued a declaration deploring the invitation to him. But this need not deter M'Bow from returning later this month, when the plans for extended cooperation will be on the agenda.

At the assembly last week, M'Bow had to field questions on the morale of UNESCO staff; on the American General Accounting Office report which attacked UNESCO spending and 70-per-cent overheads; on the political nature of a UNESCO mission to Afghanistan; and on the "new world information order". The conclusion: that the UNESCO crisis is "very deep", said an observer.

Nevertheless M'Bow may count the

meeting a success if increased cooperation between the Council and UNESCO does materialize. The 21 Council of Europe states between them pay 33 per cent of UNESCO's budget. The United States pays only a quarter. **Robert Walgate**

Australian array

Canberra

CONSTRUCTION work has begun at Culgoora, New South Wales (NSW), on the \$A32 million Australia Telescope, which, if ultimately linked with other Australian radio antennas, would possess a baseline thousands of kilometres across, making it complementary to the large radio arrays planned in the Northern Hemisphere.

The Australia Telescope will comprise two arrays of individual radiotelescopes. The Compact Array will consist of six mobile antennas each of 22 metres diameter and locatable at 37 stations along a 6-km east-west line at Culgoora. The Long Baseline Array will combine the 64-metre radiotelescope at Parkes (NSW) with one or more of the Compact Array antennas and a seventh 22-metre antenna at Siding Spring (NSW). Its 319-km baseline will permit an angular resolution of 0.02 second of arc at a frequency of 10 gigahertz, a performance that could be improved on by linkage to telescopes at Fleurs, Hobart, Alice Springs and Carnarvon, and to the 64-metre radiotelescope of the National Aeronautics and Space Administration at Tidbinbilla, which is already linked with Parkes in preparation for the 1986 Voyager fly-by of Uranus.

Approximately half the overall capital cost will be taken up by the new Cassegrain antennas, which will incorporate feed horns designed for simultaneous observation in more than one frequency band. Both arrays are due for completion by 1988, and will be operated as a national facility by the Radiophysics Division of the Commonwealth Scientific and Industrial Research Organization (CSIRO).

Jeffrey Sellar

Italy

New hope for research council

SCIENTIFIC hearts are beating faster in Italy this week, as real hopes emerge of genuine reform of Italy's major but chaotic research council, the Consiglio Nazionale delle Ricerche (CNR). Last week the Italian Parliament ended eight years of colourful Neapolitan presidency of CNR by appointing a dynamic Milanese enzymologist, Professor Luigi Rossi Bernardi, to the next four-year presidential term. With Rossi Bernardi at the helm, the control of CNR passes from the underdeveloped south of Italy to the better-organized and industrial north: certainly the most fundamental event in Italian science politics for some time.

However, Rossi Bernardi is fully aware of the magnitude of his task. "Don't promise too much", he warned last week. Rossi Bernardi knows CNR from inside as a past president of the CNR medicine and biology committee. He identifies the key issue as that of quality: the proper allocation of resources, posts and promotions on the basis of scientific merit rather than custom or seniority. This means confronting interests, inevitably a political job. But Rossi Bernardi has powerful political friends, in the socialist prime minister Bettino Craxi, in the Christian Democrat research minister Luigi Granelli (both Milanese) and in the Christian Democrat party itself (which nominated him). He belongs to no political party, and says he will manage CNR "as a scientist", with the sole objective of improving the quality of CNR-supported science; but it will certainly be no hindrance to him to have friends in the right places. It will help also, said a colleague, that he comes from a medical department, medics being "the strongest mafia — in the best sense — in the Italian universities".

Rossi Bernardi is regarded as a proven organizer, both in his earlier work at CNR and his department in Milan. He is not greatly concerned, he says, with the total budget of CNR — which will grow 10 per cent next year, roughly in line with inflation. "The real problem is how the money is spent", he says.

Rossi Bernardi must also cut through the usual trip-wires with which Italian administration likes to entangle itself, such as a law recently passed which allows CNR institutes (and other state organizations) to hold as fluid funds only four per cent of their annual budgets, the rest remaining in the state treasury. The rule was originally designed to stop misappropriation, but its effect has been catastrophic, as CNR institutes are now having to request additional funds almost weekly from Rome. "We are at a standstill all over Italy", says one Naples biologist.

Another problem facing Rossi Bernardi is the fact that something like half of all CNR scientists have published nothing for

five years (according to a survey he instituted two years ago). And the council languishes under a law which treats scien-

tists as the most routine kind of civil servant, inhibiting promotion on merit (substituting seniority) and all but forbidding appointments into CNR at high level (as laboratory directors, for example). "It's a very, very hard job", says Rossi Bernardi.

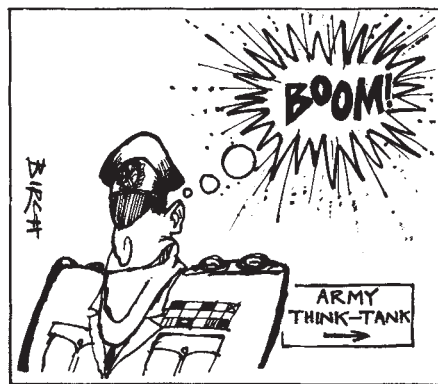
Robert Walgate

US Army research

Expelled think-tank finds Rand

Los Angeles

THE US Army's new think-tank, given the cold shoulder by faculty of the California Institute of Technology earlier this year, has at last found a home. Within the next few months, the Arroyo Center will move from its present quarters at the Jet Propulsion Laboratory (JPL, administered by Caltech) to the Rand Corporation, located in another part of greater Los Angeles.



Rand has been home to the US Air Force's think-tank since the Second World War. "We don't see that there's a conflict", said Donald Rice, Rand's president and chief executive officer, who said that "for decades" Rand has been working for both the Air Force and the office of the

Secretary of Defense as well as several other components of the Department of Defense, the Army and the Navy included.

Both the Navy and the Air Force have independent think-tanks, and the Army two years ago decided that it also needed one. The centre's charter is to examine long-term issues involving defence strategies, military operations, resource management and advanced technology as they affect the future Army.

As the centre began to set up shop last spring under Caltech's wing, the faculty balked. A majority said it was inappropriate for the university to lend its name and reputation to an organization that would lobby on behalf of military interests. Although it was acknowledged that many faculty members work on defence matters, it was argued that the university as a whole should not.

Caltech president, Marvin Goldberger, was criticized, but quickly forgiven, for having let the arrangement take place without prior faculty approval. The centre was moved to the nearby JPL campus and asked to look for another home.

At Rand, the Arroyo Center will become a fifth division joining Project Air Force and other national security research, domestic research and the Institute for Civil Justice.

Sandra Blakeslee

Nature index of biotechnology stocks

12-Month high	12-Month low	Company	Close previous month	Close 28 September	Change
14	6	Biogen (Switzerland)	8	8	0
2	1	Bio-Logicals (Canada)	1 1/4	1 1/4	0
14 3/8	6 1/2	Bio-Response (USA)	9 1/2	7 3/4	-1 3/4
14 1/8	9 3/4	Cetus (USA)	11 3/8	11	- 3/8
10 3/8	4 1/4	Collaborative Research (USA)	6 1/4	5 7/8	- 5/8
19 7/8	11 1/2	Damon (USA)	14 1/2	14 3/8	+ 1/8
26 1/4	11 3/4	Enzo-Biochem (USA)	14 1/4	14 1/2	+ 1/4
10 1/8	4 1/4	Flow General (USA)	5 3/8	5 1/4	- 1/8
42 1/4	28 7/8	Genentech (USA)	33 1/4	30 3/4	-2 1/2
10 3/4	4 1/2	Genetic Systems (USA)	6 7/8	6 1/4	- 1/8
17 1/4	8 1/4	Genex (USA)	12	9 3/4	-2 1/4
23	11	Hybritech (USA)	14 1/4	13 1/4	-1
16 1/4	7 1/4	Molecular Genetics (USA)	10 1/4	7 3/4	-2 1/2
15 1/2	8 1/4	Monoclonal Antibodies (USA)	10	10 3/4	+ 3/4
60 7/8	32 1/8	Novo Industri A/S (Denmark)	37 1/4	33 3/8	-3 7/8
22 3/4	14 1/2	Pharmacia (Sweden)	19	17 7/8	+ 1 1/8

Closing prices are for the last Friday of the month. For over-the-counter stocks, bid price is quoted; for stocks on the American and New York exchanges, the transaction price. *Nature's* weighted index of biotechnology stocks stood at 139 on 28 September, compared with 149 a month earlier. Data from E.F. Hutton, Inc.

Persistence of miracles

SIR — A leading article in *Nature* (19 July, p. 171), quite rightly questioned whether “science has nothing to say” on the subject of miracles; but it was surely wrong in attributing to science the dogma that “miracles do not occur”. In scientific laws we describe, as best we can, the pattern of precedent we observe in the sequence of natural events. While our laws do not prescribe what must happen, they do prescribe what we ought to expect on the basis of *precedent*. If by a “miracle” we mean an unprecedented event (and the resurrection of Christ is certainly described as such), then science says that miracles ought not to be expected on the basis of precedent. What science does not (and cannot) say, of course, is that the unprecedented does not (or cannot) occur.

If, as Christians have traditionally believed, the whole spatio-temporal sequence of events that make up our world owes its being to our Creator, then it is thanks to Him that our scientific explanations normally prove as reliable as they do. But by the same token, nothing whatever in our observations of “normal precedent” can make it impossible for the Creator to bring about a totally unprecedented event, if His overall purpose for His creation requires it. Christians believe that the resurrection of Christ was just such an event. We do no good to the cause of science by pretending that there are scientific grounds for denying that it could have happened.

To say this, however, is not to suggest that we should go soft on “mischievous reports of all things paranormal, from ghosts to flying saucers”, as *Nature*'s leader writer seemed to fear. For the Christian believer, baseless credulity is a sin — a disservice to the God of truth. His

belief in the resurrection does not stem from softness in his standards of evidence, but rather from the coherence with which (as he sees it) that particular unprecedented event fits into and makes sense of a great mass of data. Some of these data are historical, some derived from the cumulative experience of Christians down the ages, including himself. So his faith is not groundless. There is clearly no inconsistency in believing (with astonishment) in a unique event so well attested, while remaining unconvinced by spectacular stories of “paranormal” occurrences that lack any comparable support.

Both wishful thinking and wishful unthinking are evils, and it will not do to tumble into one for fear of the other. We cannot dogmatically exclude the ever-present possibility that the truth about our world is stranger than we have imagined.

DONALD M. MACKAY

*Department of Communications
and Neuroscience,
University of Keele,
Keele, Staffordshire ST5 5BG, UK*

SIR — Your leading article “Miracles do not happen” (*Nature* 19 July, p.171) appears to contradict itself at least twice. First you urge critical rigour for evaluating a particular miracle, the turning of water into wine by Jesus. I agree. But then you baldly state: “Miracles... do not occur — a definition by exclusion of the concept”. You apparently mean that they are excluded from all factual discourse, not just from science, since the latter would be agreeing with the very notion you wish to oppose — that science has nothing to say about miracles. Finally, you state that religious

belief “on other grounds” is unexceptionable.

But why insist on critical rigour for evaluating particular miracles, if *by definition* they do not occur? And how can you find religious belief “unexceptionable” if you deny (by definition!) some of its key tenets. It is not only the grounds of belief that are at issue here, but the content.

Your concern not to license “mischievous reports of all things paranormal” is no doubt motivated in the interest of scientific truth, but your strategy of defining away what you find unpalatable is the antithesis of scientific. It is disheartening that *Nature* should sell its empiricist birthright for the stale soup of *a priori* rationalism.

P.G.H. CLARKE

*Institut d'Anatomie,
Rue du Bugnon 9,
1011 Lausanne, Switzerland.*

SIR — Jeremy Marshall (*Nature* 30 August, p. 722) would like scientists to behave themselves and stop questioning miracles, as a sort of act of gratitude for his consent to “believe that evolution has occurred”. He objects to scientists concerning themselves with miracles because this subject “is outside the bounds of proper scientific inquiry”.

Unfortunately, miracles are sometimes invoked to explain natural phenomena on a nonscientific basis. Creationists have been particularly adept in the use of this device. Conclusions about miracles have been repeatedly reached by scientists on a scientific basis, and are not necessarily an invasion of theology. In turn, scientific inquiry into natural phenomena often reveals facts that seem sufficiently miraculous to excite admiration and awe from scientists and theologians alike.

THOMAS H. JUKES

*University of California,
Berkeley, California 94720, USA*

BioNet and GenBank

SIR — A few points in your recent news article “Limited access to BioNet” (30 August, p. 717) require clarification. The article confuses two separate projects funded by the US Government: the GenBank project, which is the national genetic sequence data bank in the United States funded by several institutes of the National Institutes of Health (NIH) and other government agencies; and BioNet, a computing resource for molecular biologists also supported by NIH.

GenBank is a contract jointly run by the Los Alamos National Laboratory and Bolt Beranek and Newman, Inc. to collect, maintain, distribute and provide computer access to nucleic acid sequences and related data. It is this work which is performed in close collaboration with the parallel effort here at EMBL in Heidelberg. The task of data collection is in fact shared between the

EMBL and GenBank groups, and data are exchanged on a regular and unrestricted basis.

BioNet is a project managed by Intelligenetics, Inc. to provide molecular biologists with computing resources, including analysis software, communications facilities, and databases accessible by computer network. (Two of the databases so provided are those produced by the GenBank and EMBL efforts.) Limitations on BioNet access are necessary because it is, at present, implemented on a single computer of finite capacity.

It should be stressed that these limitations apply to use of the BioNet computing facility, not to availability of the GenBank and EMBL databases. Both databases are available worldwide, in various forms, to anyone.

GREG H. HAMM

*European Molecular Biology Laboratory,
Postfach 10.2209,
Meyerhofstrasse 1,
6900 Heidelberg, FRG*

Nuclear power

SIR — J.R. Ravetz complains (*Nature* 9 August, p. 446) that *Nuclear Power Technology*, reviewed by Sir Alan Cottrell (*Nature* 309, 545; 1984), contains no reference to weapons technology. As deputy editor of that book I should explain that it is about nuclear power technology, not nuclear weapons technology.

While nuclear power stations and some nuclear weapons draw their energy from the same fundamental phenomenon, namely nuclear fission, the staff of the Atomic Energy Authority, by whom the material for the book was written, are not competent to write about nuclear weapons.

I am sure Mr Ravetz would not expect to see high explosives technology discussed in a book on oil-fired power stations, although the energy source is common to both.

R. H. FLOWERS

*Fuel Processing Directorate,
UK Atomic Energy Authority,
Harwell, Oxon OX11 0RA, UK*

Ethics, ethical committees and animal experimentation

from David Britt

Will the establishment of local review committees for animal experiments in Britain modify attitudes of researchers and the public?

THREE years ago, editorial comment in this journal advocated the establishment of institutional committees to scrutinize proposals for experiments on animals in a way that has proved useful in the regulation of genetic manipulation and research involving human subjects (*Nature* 293, 173-174; 1981). The committees were seen as a means of bringing together persons with differing views on animal use in an attempt to resolve practical issues of animal experimentation in a less arbitrary manner than legislation seemed to offer. Correspondence on the subject published in subsequent issues revealed a favourable reaction to the idea but this has not been translated into a rush to set up such committees in the United Kingdom. Many British research institutions have committees devoted to the supervision of laboratory animal facilities but few, if any, of these consider the ethics of proposed animal usage and are therefore avoiding the issues which the *Nature* editorial wished to see addressed.

There is a notable reluctance among British scientists to discuss the moral implications of animal experimentation, possibly because, by and large, they are uncomfortable in the area of philosophical (especially ethical) debate. As long as their work is within the law, what research scientists do is very much their own responsibility and perhaps ethical issues are a matter for individual conscience rather than open debate? The growing interest and concern of a broad section of the rational general public renders this position untenable today. Some scientists have the arrogant notion that only they can fully appreciate and weigh the issues, so that discussion with others is pointless. Other scientists and scientific administrators apparently entertain the hope that by adopting a 'low profile' approach to public concern, it will fade away. In each case the scientific community is denying any responsibility to be accountable to the public for its actions and activity.

At an individual level, is there also, perhaps, an underlying fear that too deep a questioning may reveal to the scientist the weakness of the philosophical basis on which his current, permissive position is founded?

Elsewhere in the world, scientists have been less inhibited in the debate on the ethics of using animals in experimentation and have even invited participation by the public at large. One reason often advanced for the lack of similar initiatives here is that some proponents of animal rights issues have adopted illegal, even violent, means to further their cause and thus alienated themselves from rational opinion. However, it is quite wrong to suppose that the same spectrum of personal attitudes to animal experimentation is not to be found elsewhere, even if the practical expression of extreme views antagonistic to research on animals is qualitatively different. Why this difference exists offers a fascinating area for study in social psychology, but the present situation in Britain may not be totally unrelated to the aloofness and conservatism of our scientific establishment. It is time for scientists involved in research with animals to confide in, and consult with, the general public — at least those sections of it which can be objective while maintaining a genuine concern that unnecessary suffering should not be inflicted.

If introduced to British research institutions, animal experiment review committees would assist in establishing a valuable dialogue between scientists and the lay public. Further, they would provide the scientific community with the chance to reveal its own concern for the welfare of laboratory animals and their humane, non-profligate use, thus allaying public disquiet, often fanned by inaccurate, emotional propaganda from extremist sources. Lay input might encourage scientists to be more reflective in their management and use of animals and in experimental procedures. This has happened in Sweden, for example, without discouragement of worthwhile scientific investigation.

Ethical review in Sweden

In July 1979, local ethical committees for the review of animal experiments became mandatory in Sweden. Six committees were established, one for each of the country's University regions, to cover experiments in all research institutions using laboratory animals: academic,

government and commercial. An additional committee was formed to review experiments on animals to be conducted by the military authorities, with members screened for security purposes. All the committees were made responsible to a board of the Ministry of Agriculture which still regulates the scheme and provides a secretariat and funds for essential expenses of committee members. Membership of the committees comprises equal numbers of individuals from three categories: research scientists, animal/laboratory technicians and lay persons. The sanctioning of such a high degree of participation by the general public is one of the boldest (some would argue, most reckless) aspects of the scheme. The largest committees, at Stockholm and Uppsala, have total memberships of 45, the smallest only 15. Volunteer scientist and technician members are solicited from all relevant institutions in each region; the lay members are nominated by animal welfare societies or local government agencies.

From the outset, attempts were made to ensure that the creation of an unwieldy bureaucratic monster would be avoided. Two aspects of committee organization have helped to achieve this. First, all experiments are graded (initially on a five-point scale, later simplified to two) for the degree of animal suffering likely to be imposed, and application for review is essential only for those, a minority, in which significant suffering is envisaged. In this way, the workload of the committee remains manageable. However, details of all experiments involving animals are notified to the committee, which may disagree with the grading. Second, each experimental protocol for review is discussed by the investigator and a sub-committee of only three members — one from each category — and referral of the project to consideration by the full committee is only necessary if agreement cannot be reached by the small group. Details of all projects submitted for review and sub-committee decisions are made available to all committee members. It is usual for the full committee to meet only two or three times a year.

Generally, reaction among scientists is favourable although the consensus is that

the maximum usefulness of the committees has yet to be realized. There are criticisms that the three-year trial period in Uppsala was inadequate to allow potential problems to be identified and complaints that the decision to introduce review committees on a national basis was inspired more by political expediency than by scientific or humanitarian aims.

Scientific members of the committees are recruited from among the most experienced and senior research personnel. Such individuals are frequently overburdened with meetings of committees of various kinds. Any change which threatens to increase the time to be allocated to the work of the ethical committees will deter recruitment but this may be offset as the committees grow in influence and prestige.

Technician members have been accused by anti-vivisectionist groups in the country of refraining from expressing criticism of research proposals through timidity. Where the technicians feel themselves to be in a subordinate role to the scientists there may be truth in this accusation. On the other hand, many scientists and lay members of the committees of more moderate persuasion, feel that technician representatives have a vital contribution to make through their intimate knowledge of experimental procedures and good practice in animal care, often helping the committees to frame improvements to protocols.

Lay persons volunteering for membership of the committees have a concern for animal welfare in common but differ in their attitudes to animal experimentation. Some are unable to accept that there is any justification for the use of animals in biomedical research. It is difficult to understand why these individuals elect to serve on committees responsible, ultimately, for sanctioning animal experiments — at least those which are thoughtful, well-designed and both sparing and humane in their use of animals. Because they seek nothing less than the total abolition of experiments involving animals, anti-vivisectionists must remain frustrated by the activities of the committees. If they are members of one of the committees, their philosophical position demands that they must withhold consent from any protocol, even one which may have been improved at the recommendation of fellow members to ensure greater humanity in the treatment of animals. If their wish is to disrupt research effort, this also is frustrated, because all they can hope to achieve is a prolongation of debate in committee and delay in the onset of experiments before a majority decision allows the work to proceed.

Regrettably, ethical review committees in Sweden are presently facing serious problems. It is proving difficult to obtain suitable volunteers for membership and meetings are not always well attended. This is a serious indictment of the system

because it suggests that the work of the committees is not highly valued. The problem has been exacerbated by the action of anti-vivisectionist members who have caused more and more lengthy, and circular, debate and prolonged meetings.

Proposals for radical reform of the system are currently being investigated by the Swedish Ministry of Agriculture. Included is a recommendation that each committee should be limited to a total of nine persons, the same three categories of membership to be retained but each with only three representatives. This is seen as a means of reducing administrative costs and overcoming problems of recruitment as well as facilitating the arrangement of meetings, but the idea attracts little support outside the Ministry except in so far as it promises to squeeze out participation by anti-vivisectionists. It is unlikely that this small committee could be representative of all relevant institutions within each University region and the necessary reduction in range of expertise available for consultation would also be detrimental. Because anti-vivisectionist groups wield considerable political influence in Sweden, attempts to exclude them from the committees may well be unsuccessful.

Professor Karl Johan Öbrink from the Biomedical Centre at Uppsala, one of the chief architects of the ethical review system, is also concerned about the present state of the committees and their future. He is critical of the emphasis on control which has become dominant with the codifying of the system in law and since administration of the committees became the responsibility of the Ministry of Agriculture Board. As originally conceived, the committees were to be advisory bodies encouraging consultation between investigators and other experts, with the aim of securing improvements in research design and ensuring the rational, humane use of experimental animals. In spite of misgivings about present arrangements, he remains totally convinced of the value of this kind of ethical review.

There was controversy earlier this year when a technician representative on the committee serving the Lund/Malmö region resigned after complaining that the committee system was powerless in offering protection to animals and implying that some scientists were unnecessarily callous in their use of experimental animals. This individual holds a responsible position within the University of Lund and is respected for his knowledge and experience in laboratory animal science. He is completely opposed to the views of the anti-vivisectionists, who he condemns as being largely responsible for the ineffectiveness of the committees, but, extraordinarily, finds himself advancing arguments which echo those of the opponents of animal research. He believes that the presence of anti-vivisectionists on the committees inhibits scientists and technician members from

criticizing protocols which their expert knowledge, experience or intuition persuade them are suspect, because they fear exposure of these criticisms to the media with public discredit to the investigator and institution concerned. Unsound proposals thus pass unchallenged (except, perhaps, by lay members) with the implication that unnecessary suffering may be caused to animals in these investigations. This assertion says little for the moral integrity of the scientists and technicians involved and it also seems odd that by suggesting modifications to unsatisfactory projects the committee should create unfavourable publicity for the scientific community. However, it is conceivable that insensitive media treatment of such incidents might prove damaging.

The greatest threat to the usefulness of the Swedish committees is posed by the strength of the anti-vivisectionist lobby. It is regrettable that, by striving for the unattainable, that is, immediate abolition of the use of animals for experimentation, the anti-vivisectionists are hampering the amelioration of animal suffering that is possible now. The results of this activity in Sweden have wider implications, for the bold gesture which introduced the national system of ethical review as an essential prerequisite for animal experiments has attracted international attention. By reducing the efficiency of the committees or threatening their continued existence, the Swedish anti-vivisectionists fuel critics of the scheme elsewhere who are anxious to discredit the initiative in an effort to prevent the introduction of similar schemes in their own countries.

Review in North America

In the United States scientists are currently attempting to come to terms with new guidelines on laboratory animal welfare proposed by the National Institutes of Health (NIH), the major grant-awarding body for biomedical research in the country (*NIH Guide for Grants and Contracts* Vol. 12 No. 5; 1984). One of the major proposals requires the appointment at each research institution of an Animal Research Committee (ARC). It will consider both care and use of animals, and is required to review any research proposal in which animals are likely to be exposed to severe pain or distress. Most controversial is the directive that at least one committee member must be unconnected with the institution — in effect, a representative of the general public.

The commonly held view is that practices recommended in the document could be established before the end of this year and institutions throughout the country are producing plans to ensure conformity with the requirements. Other funding agencies traditionally defer to the NIH in matters of this kind, so failure to comply could deprive a research institution of any money from external sources. The grant-awarding

bodies in the United States are the chief arbiters of what research may be carried out there and how. They have set standards which aspiring grantees have had to fulfil, including, for prospective users of experimental animals, standards of animal care and management. However, there have been a few highly publicised incidents of malpractice which suggest that the regulations require strengthening and the new guidelines are, in part at least, a response to this adverse publicity.

Some US research institutions have already, voluntarily, established committees with responsibilities similar to those outlined. A few even have lay members (often representatives of animal welfare societies) but these are exceptional. Their development has been due to the concern and enthusiasm of one or more staff members, not necessarily research scientists themselves. Ethical review was pioneered in America at Colorado State University, Fort Collins, largely through the efforts of Bernard E. Rollin, the forceful professor of philosophy, who developed a keen interest in animal rights issues and taught an ethics course to undergraduate veterinary students. The appreciation of the relevance of considering the morality of Man's use of animals which this engendered has so pervaded the University that concern for the treatment of experimental animals there quite naturally found expression in the careful consideration of experimental protocols by a committee of peers. This committee has not included participation from outside the University but plans are now in hand to change this in line with the NIH requirements.

The veterinary staff at Colorado State University gave good support to Professor Rollin's initiative. In other parts of the United States, laboratory animal veterinarians have since taken a lead in organizing similar review committees, if only to strengthen their own hands in securing improvements in maintenance and use of experimental animals. Generally, however, the veterinary profession has shown little collective concern for animals in research, preferring, as a Washington representative of the American Veterinary Medical Association (AVMA) recently remarked, "to remain neutral on the issue". There is a special AVMA committee on Animal Welfare which includes the welfare of laboratory animals in its remit and a postgraduate Diploma in Laboratory Animal Medicine has been developed which is gaining increasing acceptability as a suitable qualification for veterinarians employed in this young but expanding specialization. However, while there can be no doubt of the commitment of individual members of the profession to the cause of laboratory animal welfare, concerted effort to ensure needed improvements and reform are lacking, as if the profession is unsure of its ground in appearing to criticize colleagues

in science or medicine.

One way in which ethical concern for experimental animals is expressed is by ensuring that only competent staff are employed in their care and maintenance. Competence is achieved through experience and training, so the provision of a training course for US veterinarians responsible for experimental animal facilities is a valuable contribution to laboratory animal welfare. The appointment of veterinarians with these specialist skills is a measure of the concern of an institution for the well-being of its animals — as is the minimal or complete lack of veterinary cover found in some large research institutions in the United States and, of course, in other countries.

The training of biomedical researchers in the specific skills of caring for and using experimental animals efficiently and humanely is just as important a consideration but courses are much less formalized, or even non-existent. This deficiency is not unique to the United States.

Potential members of an animal research review committee may also benefit from some training and with remarkable alacrity the Scientists Center for Animal Welfare (SCAW) in Washington has anticipated this in a series of workshops offering help on the setting up and running of the committees along the lines of the NIH proposals. The first of three workshops planned for 1984 was held in Baltimore in May. Eighty-five of the ninety participants who completed a questionnaire after the meeting thought that committee review of research proposals for animal welfare concern was "very important". None opposed review. Most of the respondents were research scientists, veterinarians or animal care technicians. Although the group was highly selected, the survey results provide evidence of a positive attitude among American scientists and their colleagues to Animal Care and Use Committees as an important means of improving laboratory animal welfare.

The NIH recommendations offer potential for such improvement but they present several problems which institutional heads and administrators will have to consider carefully as they formulate their response. Among questions which the planners will need to ask in the context of the individual circumstances of each institution, the following are likely to be significant:

(1) Will the volume of research proposals present the committee with an excessive workload — a distinct possibility in major centres. (If only selected protocols are reviewed, will demand for public accountability be satisfied?)

(2) Will the ARC consider matters of ethical concern only or ethics and scientific merit of the projects? The two are so interconnected — animal research which is badly conceived scientifically is necessarily unethical — that consideration of both seems inevitable.

(3) Will ARC decisions be arrived at by consensus or on a majority vote? (That is, how much influence will the lone, extra-institutional member carry?)

(4) How will the ARC confirm that investigators comply with its decisions? (Will there be any follow-up, reporting back, committee inspections or monitoring procedures?)

(5) Will there be any attempt to standardize ARC policy and criteria for decision making between institutions?

(6) What will be the legal status of ARCs?

(7) Will investigators have a right of appeal against decisions of the ARC?

(8) Concerning the non-institutional member: what kind of person should be selected? What should be the terms of the appointment? Should he/she be remunerated? How can the need for confidentiality be reconciled with conscience to protect (i) the interests of the institution and (ii) the integrity of the lay member.

The current proposals allow research institutions to select the outside member(s), but they will want to appoint persons from the community with sufficient objectivity to be credible as representatives of public opinion.

Several of the problems listed above have already been faced in Sweden and in Canada, where institutional Animal Care Committees have been in existence for several years. In Canada, such committees are a feature of all institutions in which experimental animals are used. Their responsibilities include the sanctioning of all work involving animals (in teaching, field studies, research and trials), ensuring high standards of animal care and the review of research protocols. In Ontario and Alberta, the committees are mandated through provincial legislation. The Canadian Council on Animal Care (CCAC), an authoritative body within the Association of Universities and Colleges in Canada, has recommended terms of reference and guidelines to which the committees conform. Periodically, site visitors are arranged by CCAC to assess the efficiency of each institution's committee and help effect standardization on a federal basis. It is recommended that membership should include lay persons to represent community interests; lay person also serve on assessment panels for site visits. Existing lay members of Animal Care Committees represent major animal welfare societies or are members of professions: law, education, the church. The committees should be primarily concerned with the ethical aspects of the suggested procedures and the property of suggested methodologies, since grant-awarding agencies use peer review to confirm the scientific merits of projects seeking financial support.

The Canadian Federation of Humane Societies, representing 36 member organizations, and somewhat akin to Britain's RSPCA, provides an effective mouthpiece

for public animal welfare concerns in Canada. Its view is pragmatic on the issue of using animals in experiments, but there is a growing tendency, within the Federation, to question the continuing need for their use in biomedical research. It is fully supportive of the institutional Animal Care Committees and distributes publications to guide their members.

There is no shortage of animal welfare groups in the United States, one or more of which might adopt a similar role to that of the Canadian Council on Animal Care in providing oversight of the new committees when established, perhaps in conjunction with officers of NIH. Support similar to that provided by the Canadian Federation of Humane Societies would help to ensure the effectiveness of the American committees.

Lessons for Britain?

How relevant is all this to the situation in Britain? It is naive to suppose that a regulatory or advisory system which proves valuable in one country can be transposed to another with similar expectations. Equally, it is foolish to ignore interesting developments in other parts of the world which are relevant to one's own situation. Successes and failures of overseas initiatives can inform and guide the planners of suitable indigenous approaches. This is not to deny that there are major historical, cultural, legislative and logistical differences between the countries included in this discussion which create formidable barriers to the easy exchange of ways of tackling common problems.

One difference which is frequently advanced to suggest that Britain has nothing to learn from groups overseas in respect of animal welfare, is that legislation for the protection of animals in Britain pre-dates that of any other modern State. The importance of this tends to be over-emphasized and other nations have not been so tardy as is sometimes supposed. Sweden's first animal protection act, for example, dates back to 1944 and it is arguably more relevant to the present day than the British legislation formulated in 1876, no matter how amenable to reinterpretation that law may have proved.

Because of this misconception — that all other countries are 'starting from scratch' in animal protection — regulations which the latter introduce are interpreted as being nothing more than attempts to bring themselves 'up to our level'. In fact, the practical and philosophical expression of concern for improvements in laboratory animal welfare in countries like Sweden and Canada has outstripped legislative provision here — even if the ethical awareness of some scientists and others involved with laboratory animals in Britain remains as keen as that of their counterparts anywhere in the world.

The present British regulatory system is

applauded as a model of rationality and efficiency by some and described as wholly inadequate by others. The latter group includes not only the hard-line opponents of animal experimentation but scientists working within the system. Wherever the truth lies, future Government legislation, based on an appraisal of the current system, promises to significantly tighten controls. But, no matter how stringently the law may be drawn and no matter how rigidly it may be enforced, at the interface between the scientist and his subject, only the attitudes of the scientist are important. A heightened awareness by experimenters of the responsibilities associated with the use of animals and increased sensitivity to their needs and their capacity to suffer will achieve much more for laboratory animal welfare than new legislative measures (although legislation to encourage the introduction of institutional review bodies here to help achieve these desirable changes would seem appropriate!). These were the kind of thoughts which prompted the Uppsala experiment and which make the Fort Collins initiative so important, where the ethos of these institutions has developed to permit the expression of a rational, caring approach to the use of animals in research.

The committees in Sweden and North America (existing and planned) are answerable directly to the institution chief or administrative head and contain very senior staff as members. As a tangible benefit to the institution, they provide an important public relations service by ensuring compliance by institution personnel with existing national and local regulations on the care and the use of experimental animals. Beyond this, they provide a guarantee that only 'clean' work is undertaken, this assessment having been reached not merely by the scientists alone. The attraction of such guarantees to institution chiefs is obvious and, equally, it is important to the prestige of the committees that they are seen to stem from the highest institutional authority. The same recognition will be critical to the success of any similar committees which may be developed in Britain.

We can profit from the problems which have occurred in Sweden by avoiding some of the pitfalls. The anti-vivisection movement in Britain is proportionally smaller than its Swedish counterpart. The Nordic Society Against Painful Experiments on Animals (Nordiska Samfundet mot Plågsamma Djurförsök, NSPD) has a membership of 38,000 in a country with a total population of only 8.5 million. Furthermore, so evenly divided is the country politically that 70,000 votes could change the government. In such a situation, any sizeable pressure group can exert considerable political influence and this power is used astutely by NSPD. To deny representation of minority, extremist groups opposed to animal experiments on

ethical committees in Britain would not be thought tendentious. Here convinced opponents of animal research will not be favourably impressed by the introduction of research review procedures. Certainly for the law-breakers, the situation has gone beyond the stage at which participation in this sort of constructive dialogue is possible.

In Sweden and America there is still opportunity for discussion and compromise between groups of all shades of opinion. Once this ground is lost, it is unlikely to be regained — which is why attempts to exclude Swedish anti-vivisectionists from the ethical committees may have retrograde effects.

Several factors seem to favour the introduction of local ethical committees to Britain and, if forthcoming legislation follows the recommendations of the 1983 Government White Paper, *Scientific Procedures on Living Animals* (HMSO, cmd 8883), this will strengthen the case for their establishment, in defining a role for them in careful scrutiny of applications for project licences before submission to the Home Office.

We have available through the Home Office Inspectorate, an excellent means of supervising institute-based ethical committees and ensuring a desirable measure of inter-institutional standardization. By close liaison with the committees, inspectors could exert their own influence more effectively in the regions. Undoubtedly, the collaboration and support of the Inspectorate will be vital to the successful functioning of the committees.

Major animal welfare groups in Britain favour the concept of local ethical committees and can be relied upon to participate in any initiatives to establish these bodies — providing that they are more than window dressing for research institutions.

Perhaps an Uppsala-type experiment is necessary here before British scientists will be persuaded of the value of the animal research review process in the context of University activities. This will be considered more fully at a meeting being convened in the University of Liverpool on 1st and 2nd October in response to a demand from scientists for improvements and alternatives to animal experimentation.

Scientists will groan at the prospect of 'yet another committee', but if the protestations of genuine concern for animal welfare which are frequently voiced by animal experimenters mean anything, then priority must be given to this kind of provision. □

David P. Britt was awarded a Churchill Trust Travelling Fellowship (1984) to study ethical monitoring of animal experiments in North America and Sweden. He is at the Department of Veterinary Parasitology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK.

Thunder and electric storms

A neat measurement of the stratospheric electrostatic fields throws light on the link between solar activity and terrestrial climate.

CAREFUL readers of the scientific literature are forgivably suspicious of articles that refer to the very distant past. The sceptical will ask whether the authors of such an article have sinister motives for ignoring recent work, and will probably dismiss the possibility that they have managed to resuscitate an important problem whose trail runs cold before, say, 1980. So what is to be made of the contribution now published by R.H. Holzworth and T. Onsager of the University of Washington (Seattle) and P. Kintner and S. Powell of Cornell University to the problem of atmospheric electrostatic fields? Their paper (*Phys. Rev. Lett.* **53**, 1398; 1984) manages one reference to a regular paper (by Holzworth) after 1980, but is mostly spattered with references dating from the 1920s and earlier.

Holzworth *et al.* have nevertheless managed to break new ground, and in a stimulating fashion. What they have done is to equip two balloons so as to measure the electrostatic field in the Earth's atmosphere at an altitude of 26 km. The balloons were released from New Zealand in March last year, and allowed to drift for two weeks eastwards across the Pacific at a latitude of 45°S. The lateral separation between the balloons was never less than 750 km. The objective was to compare the electric field strength at two such widely separate locations. The conclusion that the two records are very similar, but that fluctuations of field strength from day to day are much greater than expected, have an important bearing on the understanding of atmospheric electricity.

One of the obvious early references in Holzworth *et al.* is to C.T.R. Wilson, whose design of the closed chambers that became the workhorses of early cosmic-ray measurements was originally part of an investigation of atmospheric electricity. What Wilson knew was simple — that thunderstorms consistently carry positive electric current upwards in the atmosphere and that near the surface of the Earth, the electric field strength (the gradient of electric potential), while variable, can amount to several hundred volts per metre. Wilson had the wit to recognize that the cumulative effect of thunderstorms must be to some degree offset by a less spectacular but continuous downward electric current, for the electrical potential between the surface of the Earth (negatively charged) and a presumed upper conducting layer (the ionosphere) cannot

increase indefinitely.

But how can the atmosphere conduct electricity, even at the low rate suggested by the observations? Wilson was the first to show that the production of ions by cosmic rays is the essential mechanism for rendering the bulk of the atmosphere conducting. The potential difference between the surface of the Earth and the ionosphere is typically no more than some tens of thousands of volts — the large potential gradient near the surface quickly trails off with increasing altitude as cosmic-ray ionization increases the conductivity of the atmosphere — at 26 km Holzworth *et al.* measure fields of less than 1 V m⁻¹.

The immediate significance of the new balloon measurements is that they have demonstrated a technique for measuring electrostatic fields beneath the ionosphere that is bound to become established if sparsely used. From this first batch of data, the conductivity at 26 km seems virtually constant for days on end, while the electrostatic fields measured by the two balloons, even when separated by up to 3,000 km, are identical to within 10 per cent for most of the time (perhaps 70 per cent of the 16 days of common observation). That is understandable if, indeed, the field at 26 km can be likened to the electrostatic field between two parallel conducting plates, one the ionosphere and the other the ocean surface.

For then, given the high conductivity of the upper and lower surfaces, the relaxation of the distribution of charge must be rapid. The local effects of thunderstorms carrying positive charge into the stratosphere would be quickly dissipated. Holzworth *et al.* point out that there nevertheless seems to be a daily variation of field strength at 26 km, at least when 10-minute records for consecutive days are averaged together, with a possibly shallow maximum of field strength at 20h Universal Time. The records, however, show no drift in the timing of the diurnal variation with the eastward passage of the two balloons across the Pacific, which provisionally gives the lie to theories that the global pattern of the vertical electrostatic field is fixed in relation to the position of the land masses on the surface of the Earth.

The chief interest of these data, according to their authors, is the enormous variation in the pattern of the vertical electrostatic fields measured by each balloon on different days. Field strengths are generally less at local morning than

after noon, but otherwise the variability is so great that one day's record is unrecognizably different from the next's. The inference, which is important, is that there must be substantial fluctuations (perhaps by a factor of two) in the global vertical current in the atmosphere, and on time-scales of the order of 10 minutes, perhaps themselves the consequences of fluctuations of thunderstorm activity near the surface of the Earth.

While Holzworth (among others) has previously been responsible for vertical field measurements at altitudes up to 30 km, the twin balloon measurement seems to be the first of its kind. Evidently the conclusion that there are large short-term fluctuations of the global electrostatic field could not have been sustained by measurement with a single balloon (for local thunderstorms might always have been responsible). One obvious benefit that lies ahead is a more detailed correlation of the global field fluctuation with observed thunderstorm activity (presumably anywhere in the world). The point is obviously important because, while there seems little dispute about the correlation between thunderstorm activity and the strength of the electric field, the degree to which the former may be excited by the latter, involving as it does the non-linear dynamics of condensation in cumulus clouds, remains to be settled.

So does the possibility that the electrostatic field beneath the ionosphere may provide a physical link between sunspot activity and terrestrial climate. Ralph Markson has argued that solar activity, by affecting the cosmic-ray flux and thus the conductivity of the atmosphere, might influence the occurrence of thunderstorms (*Nature* **273**, 103; 1978), but by a mechanism which has been disputed (L.C. Hale, *Nature* **278**, 373; 1979) and which again hangs on the degree to which electric fields and thunderstorms are mutually stimulating.

Unfortunately, given the short-term global fluctuations of the vertical field, it is by no means clear how quickly careful monitoring of the electrostatic field by an *in situ* measurement will be feasible, and how quickly that will yield direct evidence bearing on the question whether solar activity and terrestrial weather are in any way linked by atmospheric electric fields. At least another sunspot cycle will go by before matters are cleared up.

John Maddox

Cosmology

Cosmologists in the dark

from Joseph Silk

In their continuing quest to unravel the true nature of the Universe, cosmologists are beginning to pursue a most bizarre trail. Never mind that the Universe contains more mysteries than a whole galaxy of astronomers could ever unravel. Now we are learning that all the clues may be heralding a false dawn. What we see may have little to do with the true distribution of matter in the Universe. If galaxies formed only in specially-selected regions, then much of our painstakingly-acquired knowledge of the large-scale structure of the Universe, based on decades of observations with the largest telescopes available, may as well be discarded.

Could the Universe really be fooling us so effectively? The search for the holy grail of elementary particle physics, an ultimate theory of the unification of the fundamental forces of nature, has provided a persuasive argument. One of the successful outcomes of this approach has been the development of grand unification theories (GUTs) which incorporate the electromagnetic and nuclear (weak and strong) forces. Many GUTs exist, and some have a property of profound cosmological importance. There is likely to be a phase transition, associated with the spontaneous breaking of symmetry between strong and electroweak forces, in which sufficient entropy is liberated to trigger an epoch of exponential expansion. This inflation of the Universe can explain several outstanding cosmological conundrums, including the isotropy of the Universe and the origin of the fluctuations from which galaxies formed. One of its most remarkable predictions, however, is that the Universe should be, almost precisely, flat. If spatial curvature is negligible, the Universe will expand at a rate that barely overcomes its self-gravitational attraction. It will continue to expand, slowly decelerating, into the remote future.

In such a situation, simple application of Einstein's or even Newton's theory of gravity, if suitably generalized, to the expanding Universe leads one to infer a critical density of matter that our observed Universe must contain. This is equal to $3H_0^2/8\pi G$, where G is Newton's constant of gravity and H_0 is Hubble's constant for the linear expansion of the Universe, and amounts, very roughly, to an average density of about three atoms per cubic metre.

To compare what we actually see with this prediction, it is convenient to define the cosmological density parameter Ω to be the ratio of the actual density to its critical value, above which the Universe would eventually recollapse. Inflation predicts that $\Omega=1$ to within 1 part in 10^4 . However,

astronomers are almost unanimous in their assertion that the density of all observed matter amounts to $\Omega \approx 0.1$, with a factor of two or so uncertainty. The loophole for reconciling the observed Universe with the inflationary theory is that a component, even a dominant one, of the mass density that is uniform over the scale of galaxy clusters or superclusters would be very difficult to detect. Certainly, such a smooth backdrop has no direct influence on the dynamics of galaxies or those of galaxy clusters, which provide the principal means for determining Ω .

The evolution of the observed large-scale structure does depend very sensitively, however, on the distribution of an underlying dark component of matter. Gravitational instability, by which small density fluctuations in the early Universe have amplified into galaxies and galaxy clusters, would have been quenched had the dominant mass distribution always been, for example, very hot and consequently very uniform. On the other hand, if astronomers' measurements of $\Omega \approx 0.1$ are taken literally, one encounters a severe problem in understanding the extreme isotropy of the cosmic microwave background radiation. Recently reported by Juan Uson and David Wilkinson of Princeton University to be isotropic to 3 parts in 10^5 on an angular scale of 4.5 arc min, the background radiation should sample the seed fluctuations in the matter distribution at the epoch when the radiation was last being scattered. Lowering Ω reduces the effectiveness of gravity in amplifying these fluctuations, and to obtain the observed structures, one would infer for the most plausible models correspondingly larger ripples in the microwave background than are seen. With $\Omega=1$, this problem need not arise.

A workshop devoted to increasing our understanding of the evolution of the large-scale structure of the Universe, and lasting eight months, has just ended at the Institute for Theoretical Physics in Santa Barbara, California. One of the most significant outcomes was a scheme for reconciling the dynamic measurements of Ω with a dense universe. The idea, developed most fully by J. Bardeen (University of Washington) and N. Kaiser (University of California, Berkeley), is that galaxy formation occurs only in a biased subset of the real Universe, at points where the critical density is particularly high. Consider a random field of density fluctuations (a three-dimensional analogue of the pattern of waves on the ocean surface) with power on both large and small scales. Suppose that the bright galaxies form only at the peaks of the density fluctuations. It

turns out that the large-scale structure of the galaxy distribution can be much more pronounced than that of the underlying distribution. One way to visualize this is to imagine filtering noise with a low band-pass filter: the surviving intensity peaks tend to be more strongly correlated than the unfiltered distribution. A similar mechanism has been applied by Kaiser to explain the anomalously large clustering of rich clusters of galaxies.

A precise physical mechanism for producing such a bias has not so far been worked out. It is likely to involve feedback of energy by the first galaxies to form. Since the largest peaks on any given scale collapse first, one could readily imagine that some form of negative feedback, suppressing later formation of galaxies, might lead to a plausible biasing mechanism. A scheme of biased galaxy formation resolves many of the outstanding problems of the large-scale structure of the Universe. The weakly correlated dark matter leads to predictions of low galaxy peculiar velocities, in accord with astronomical measurements, and the spatial correlations predicted for the galaxies resemble what is observed.

As for galaxies, galaxy groups and clusters, the presence of gravitationally dominant dark matter within their luminous confines seems to be an observational fact. A second theme to emerge from the Santa Barbara workshop and one that is reviewed on page 517 of this issue by Blumenthal *et al.* is that cold dark matter gives rather a good fit to the prescription for our Universe. This dark matter, which consists of hypothesized weakly-interacting elementary particles such as photinos or axions, was sufficiently cold in the early Universe that fluctuations developed by gravitational instability over a wide range of scales, down to and even below that of a typical galaxy. The massive neutrino is the prototype of a species of weakly-interacting particles that would have been sufficiently hot, when freely streaming at the speed of light, to have erased all substructure on scales below those of great galaxy clusters. A hot dark matter scenario for galaxy formation has several unpalatable features, most of which arise from the constraint imposed by the near uniformity of the Universe on large scales. This model guarantees that the first scales to have undergone collapse and allowed galaxy formation to occur could only have done so at what seems to be an unacceptably recent epoch. Cold particle models, on the other hand, undergo hierarchical clustering at a relatively early stage, small scales condensing first, followed by successively larger-mass scales.

The cold particle-dominated universe appears to provide a satisfactory understanding of the distribution of dark matter on scales from those of dwarf galaxies to those of the great clusters of galaxies. Combined with a biasing mechanism, it reconciles the inflationary prediction of a

universe at critical density with the astronomers' Universe. Even the distribution of fluctuations predicted by the inflationary theory is, according to Blumenthal *et al.*, reflected in the characteristic distribution of galaxy masses and densities. What remains to be shown is whether the giant voids in the galaxy distribution, such as that in Bootes, and the enormous aggregates of galaxies, such as that in the Serpens-Virgo region some fifty million megaparsecs across, can be

reconciled with such a theory. One of the most tantalizing aspects of the currently moribund idea of a universe dominated by massive neutrinos was that it naturally predicted the presence of large filaments and sheets of galaxies, separated by huge voids. It may not be so easy for a cold particle-dominated universe to meet the same challenge. □

Joseph Silk is in the Astronomy Department, University of California, Berkeley, California 94720.

Astrophysics

Oscillations of degenerate stars

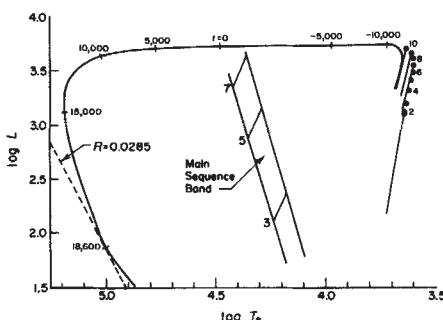
from Arthur N. Cox

RECENTLY there has been renewed interest in the oscillations of degenerate dwarf stars. Very-low-luminosity variations were first seen in the white dwarf HL Tau-76 by Landolt some 20 years ago¹, and since then, luminosity variations with periods ranging from hundreds to thousands of seconds have been detected in many other white dwarfs. The past few months have seen several predictions confirmed by new observational discoveries, and the scope of our research and understanding of degenerate star oscillations greatly expanded.

The stars that have stimulated the interest are very small, with radii ranging from 1–10 per cent of the radius of our Sun, only slightly bigger than that of the Earth, and a probable mass of 0.6 solar masses ($M_{\odot}$). Their mean density is therefore very high, about 10^6 g cm^{-3} , implying the presence of electron degenerate matter throughout most of their mass. Because degenerate matter conducts heat easily, white dwarf stars have core temperatures as low as 10^7 K or less and, thus, are much cooler than any other stars except the lower-mass very-faint red dwarfs. White dwarfs produce no nuclear energy because any potential fuels, such as hydrogen and helium, are found only in their cool surface layers, so their luminosity is low. Because of their very small size and low luminosity (less than 1 per cent of that of our Sun) it has taken billions of years for the surfaces of these stars to cool to their current temperatures, typically 10,000 K.

Stellar evolution theory predicts that stars with masses of up to at least 6 $M_{\odot}$ evolve to become white dwarfs (see the figure). As these stars burn their initially copious supply of hydrogen to helium and the helium to carbon, oxygen and neon in a growing electron-degenerate core, they expand and increase in luminosity, to become red giants or supergiants. When they have reached a radius of perhaps 100 or more solar radii, they lose mass, sometimes to produce planetary nebula shells, by shedding all but the CNO core resulting from the hydrogen and helium

burning. The remaining core has a surface temperature typically of 150,000 K and a luminosity of 10,000 solar luminosities. Because the stars have lost the layers in which nuclear processes take place, they can no longer produce any new energy to sustain their luminosity and they rapidly (within no more than 10,000–100,000 years) become very small and hot. Finally,



Hertzsprung-Russell diagram where luminosity is plotted versus the surface temperature, showing the evolution of a 0.6 $M_{\odot}$ star (adapted from ref.10). The initial increasing luminosity tracks for stars of 3, 5, and 7 $M_{\odot}$ are indicated as the main sequence band. The final stages of nuclear burning of the 0.6 $M_{\odot}$ star result in 10 helium shell flashes where the luminosity is greatly increased for durations of hundreds of years. These are marked at the very coolest surface temperatures where the mean emitted luminosity is still increasing with time. Zero time is arbitrarily set at the moment when the surface temperature increases to 30,000 K. The track represents hot dwarfs (upper left), turning as they slowly cool to become very faint white dwarfs (below the bottom of the figure). The minimum radius of the hot dwarfs is 0.0285 solar radius, and later this radius decreases as the conventional white dwarf is formed.

slow cooling turns these hot dwarfs into faint white dwarfs after millions of years.

Many hot luminous stars have been discovered and very recently five of them have been found to show 5–30 minute period variations in their luminosity of a few per cent (ref.2 and H.E. Bond & A.D. Grauer, personal communication). These stars constitute a new class of variable, called PG variables, after the first studied PG1159-035 which is perhaps the hottest

variable star in the sky. Oscillations of these stars are thought to result from the periodic ionization of carbon and oxygen in their surface layers. While they oscillate, the stars deviate from their spherical shape, hence their classification as nonradial pulsators. Last year, Starrfield *et al.* predicted that the stars would pulsate in nonradial g modes only if about half of the mass of their outer pulsation-driving layers was oxygen, and helium comprised only a small fraction³. Now, a large oxygen abundance has been verified spectroscopically⁴. Observations of helium at the stars' surfaces, however, indicate incomplete loss of this fuel. If there is enough helium, another shell flash may occur during the stars' evolution. Further studies of the longer-period K1-16 planetary nebula central star listed in the Kohoutek catalogue are underway to see what pulsations would be predicted at earlier and cooler stages of this luminous stellar evolution.

The same cyclical ionization of hydrogen and helium that drives the pulsations of Cepheids also causes the much cooler and older ZZ Ceti stars, the classical white dwarfs, to pulsate. The α and γ effects drive motions by both blocking and hiding the radiative luminosity at key compressed phases of the pulsation cycle. At subsequent phases, the radiation is allowed to flow more strongly so that the oscillations are driven as the stars re-expand.

A further intriguing class of variable stars is the DB white dwarfs. These stars have helium-rich atmospheres and, probably, helium-rich pulsation-driving layers with temperatures of 100,000 K or more. They do not have hydrogen ionization driving, but they do have the driving mechanisms from the ionization of helium. Winget *et al.* predicted that these helium stars should pulsate in nonradial modes⁵, and recently this has been confirmed in three cases (ref.6 and H.E. Bond & A.D. Grauer, personal communication).

What does the future hold in store? Several groups⁷⁻⁹ have predicted that radial pulsations in high overtones should occur in white dwarfs. Perhaps the next observational discovery will be that these approximately 1-second oscillations also exist. □

1. Landolt, A.U. *Astrophys. J.* **153**, 151 (1968).
2. Bond, H.E., Grauer, A.D., Green, R.F. & Liebert, J.W. *Astrophys. J.* **279**, 751 (1984).
3. Starrfield, S.G., Cox, A.N., Hodson, S.W. & Pesnell, W.D. *Astrophys. J. Lett.* **268**, L27 (1983).
4. Sion, E.M., Liebert, J. & Starrfield, S.G. *Astrophys. J.* (in the press).
5. Winget, D.E., Van Horn, H.M., Tassoul, M., Hansen, C.J. & Fontaine, G. *Astrophys. J. Lett.* **268**, L33 (1983).
6. Winget, D.E., Robinson, E.L., Nather, R.E. & Balachandran, S. *Astrophys. J. Lett.* **279**, L15 (1984).
7. Cox, A.N., Hodson, S.W. & Starrfield, S.G. *Lect. Notes Phys.* **125**, 458 (1980).
8. Saio, H., Winget, D.E. & Robinson, E.L. *Astrophys. J.* **265**, 982 (1983).
9. Starrfield, S.G., Cox, A.N., Hodson, S.W. & Clancy, S.P. *Astrophys. J.* **269**, 645 (1983).
10. Iben, I. & Renzini, A. *A. Rev. Astron. Astrophys.* **21**, 271 (1983).

Arthur N. Cox is a Fellow of the Los Alamos National Laboratory, PO Box 1663, Los Alamos, New Mexico 87545.

Vaccines

In search of novel immunogens

from Peter Newmark

To judge by a recent meeting*, empiricism and tradition still offer at least as much to the production of new vaccines as rationalism and peptides. But the possibility of basing new vaccines on synthetic peptides that correspond to immunogenic parts of a pathogen's structure is acting as a great stimulus to immunochemists and their allies.

The initial problem with the peptide approach to vaccines is to identify immunogenic peptides. Were it possible to do that from no more than the amino acid sequence of a protein, life would be easy. But, although it is possible to predict crudely which segments of a sequence will form the surface of a folded protein, and although such segments, in peptide form, are likely to be antigens, they will not necessarily be immunogenic, in the sense of evoking a sufficient immune response to neutralise the pathogen. Moreover, it is inevitable that many, and probably most, immunogenic sites on a protein are not composed of contiguous amino acids but rather of amino acids that are juxtaposed by protein folding and so cannot readily be mimicked by a synthetic peptide.

For the most part, immunogenic peptides have been identified by screening candidates with monoclonal antibodies that neutralize the relevant pathogen *in vitro*. A promising new technique, requiring neither prior peptide synthesis nor sequence information was described by J. Nunberg (Cetus Corporation). The cloned gene of an immunogenic protein of a pathogen (feline leukaemia virus in the first case) is cut into random short fragments by DNase I. The fragments are cloned into Lambda phage 'Charon 1b' in such a way that they are expressed. Phage colonies are screened with a monoclonal antibody that neutralises the pathogen. Those that react with the antibody must be expressing an antigenic peptide from an inserted gene fragment which is then sequenced (see *Proc. natl. Acad. Sci. U.S.A.* **81**, 3675; 1984). In that way, Nunberg reported, it had been possible to identify a 14-amino acid immunogen of the envelope protein of feline leukaemia virus. The corresponding synthetic peptide was able to compete with the virus for antibody. Moreover, when attached to a carrier and used to vaccinate guinea pigs, the peptide elicited a partial immunogenic response which was markedly enhanced by subsequent administration of a low dose of killed virus. An efficacy trial of this approach to a vaccine against feline leukaemia virus in cats is now being set up.

E. A. Emini (Merck Sharp and Dohme) reported a roundabout, but revealing, identification of immunogenic sites on the hepatitis A virus, a relative of poliovirus. Starting from the fact that antibodies against hepatitis A virus demonstrate some cross-reactivity with peptides representing major immunogenic sites on the VP1 protein of poliovirus type 1, and taking into account the observation that these peptides could prime animals to respond to a sub-immunogenic inoculum of hepatitis A virus, Emini expected to find that certain sequences within the VP1 of hepatitis A virus closely matched those of the poliovirus peptides. But the relevant cDNA sequences (J. Hughes, Merck Sharp and Dohme; B. Baroudy, National Institutes of Health, Bethesda) revealed no such match. Nor was there a concordance in the calculated secondary structures of the two VP1s. Only when a comparative analysis of the predicted surface features of the two proteins was made, was a striking similarity observed. Based on this comparison of 'shape', Emini *et al.* have synthesized a hepatitis A virus peptide, monoclonal antibodies against which have virus neutralising activity, mostly cross-react with the virus and in some cases compete with antibodies in the sera of humans infected with hepatitis A virus. These data underline the importance of the 'shape' of the surface of immunogenic proteins and explain why much effort is being invested into an understanding of 'shape' and its determinants.

Another way to investigate the relationship between structure and immunogenicity is to document any differences in sequence between a virus that can be neutralised by a monoclonal antibody and any genetic variants that resist neutralization. The expectation would be that resistance is the result of mutations within the sequences that encode antibody-binding regions of the protein. However, D. Diamond (State University of New York, Stony Brook) reported that expectations were confounded in the case of poliovirus type 1. Two regions within its VP1 protein have been identified as antibody binding sites but virus variants that resist neutralization have amino acid changes elsewhere on VP1 or even on VP2 or VP3 proteins. Presumably, within the virus proper, the mutated amino acids are either juxtaposed to the identified antibody binding sites or affect conformation at a distance.

Also intriguing were the sequence comparisons of an attenuated strain of poliovirus type 3 (used in the Sabin-type vaccine), its parent and some strains that are revertant in terms of their neurovirulence (J. Almond, Leicester University). The attenuated strain had ten base changes

from its parent, only three of which give rise to amino acid changes. Contrary to expectation, none of these three is consistently of the parental type in the revertants; the only consistent base reversion is in a non-coding position. By recovering the poliovirus from sequential nappies of a vaccinated child the same base reversion has been documented to occur within 36 h of vaccination with the attenuated strain. Clearly the reversion is strongly selected for — but why?

A whole session of the meeting was devoted to the possible resurrection of vaccination with vaccinia virus, using it to carry and express a wide variety of immunogens. This possibility was recently put in its proper perspective in these columns by J. Beale (6 September p.12), and F. Fenner (John Curtin School of Medical Research, Canberra) provided more perspective at the meeting. Both commented that the known, though very infrequent, complications of smallpox vaccination could be a barrier to the return of vaccinia inoculation. But R. Buller (National Institutes of Health, Bethesda) described some parameters by which the virulence of vaccinia virus can be reduced if the thymidine kinase gene of the virus is used as the target for insertion of new immunogen genes.

In a novel application of the vaccinia approach H. Koprowski (Wistar Institute) reported that vaccinia virus recombinants containing the gene for the immunogenic glycoprotein of rabies virus are able to protect experimental animals against an intracerebral challenge by the virus. K. Cremer (National Institutes of Health, Bethesda) reported that intradermal inoculation of mice with vaccinia virus carrying the gene for glycoprotein D of herpes simplex virus type 1 provided protection against both type 1 and type 2 virus and also inhibited latent ganglionic infection by type 1 virus. And E. Paoletti (New York State Department of Health, Albany) described inoculation experiments using a vaccinia virus in which the genes for immunogens of hepatitis B, herpes simplex and influenza viruses had been inserted at different sites; each provoked an antibody response, demonstrating the potential of vaccinia virus as a polyvalent vaccine.

Another approach to new polyvalent vaccines was described by P. Valenzuela (Chiron Research Laboratories). The idea is to recombine the gene of a second immunogen with that of hepatitis B surface antigen (HBsAg) so that the second immunogen is assembled into, and presented on, the HBsAg particles that engineered yeast and mammalian cells produce. As an initial step in that direction, the gene for glycoprotein D of herpes simplex virus has been recombined with that of HBsAg and a lysate of yeast containing the recombinant has been shown to react with antibodies against both proteins.

Ironically perhaps, the recombinant was made by inserting the herpes gene into the

* 'Modern Approaches to Vaccines', Cold Spring Harbor, 12–16 September, arranged by R. Chanock and R. Lerner. The proceedings will be published in January 1985.

'pre-S' sequence that precedes the 'S' sequence of HBsAg and encodes a 55 amino acid extension which is poorly represented in serum HBsAg particles and is not found in yeast HBsAg particles because only the S gene has been engineered into yeast. The irony is that the pre-S encoded extension may be too important to be ignored or disrupted according to S. Kent (California Institute of Technology). Data he presented suggest that the pre-S extension is more immunogenic than the S gene product; that it should be included in hepatitis B vaccines; that a synthetic peptide corresponding to the amino terminus of pre-S might itself be the basis of a vaccine; and that antibodies against the peptide will find diagnostic uses. Note, however, that

serum and yeast HBsAg particles have been shown to protect chimpanzees against hepatitis B virus challenge, as has vaccinia virus carrying the S gene (see J. Beale *op.cit.*)

Space forbids mention of many other diseases for which a modern (sometimes not so modern) approach to a vaccine was reported. It should, however, be recorded that a session and a half was devoted to the prospects of a malaria vaccine and to the repetitive sequences within most of the potential malaria immunogens; the bulk of what was reported has recently been discussed in these columns (27 September, p.300 and 16 August, p.541).

Peter Newmark is Deputy Editor of Nature.

Biological aerodynamics

Flight of fancy planned for the largest pterosaur

from Kevin Padian

AT one time, pterosaurs, the 'flying reptiles' of the Mesozoic era, seemed relatively simple to understand. They were considered to be bat-like inferior prototypes of birds, incapable of flapping flight or of efficient locomotion on land, no more than an early experiment in vertebrate flight. The giant *Pteranodon*, with its 7 m wingspan, seemed to be at the upper limit of any flying creature's size. Their replacement by the 'real masters' of the air, the birds and the bats, could easily be explained by a factor as trivial as a change in the average wind speed at the end of the Cretaceous (Bramwell, G.R. & Whitfield, C.D. *Phil. Trans. R. Soc. B* **267**, 503; 1974).

Recent work has called for a re-evaluation of these bizarre animals; none more so than the discovery of the giant *Quetzalcoatlus northropi* (Lawson, D.A. *Science* **187**, 947; 1975) which, with a wingspan currently estimated at 11 m, was 50 per cent larger than *Pteranodon*. This remarkable beast was responsible for a recent gathering* of over 30 engineers, aerodynamicists, palaeontologists and inventors in Pasadena, California, to discuss the possibility of building an accurate radio-controlled flying model of the creature. The purpose of such a venture would be not only to challenge the potential of ornithopter technology, but also to use the problems posed to the aerodynamicist to illuminate the biological limitations of the largest animal ever known to fly.

Before any accurate flying model of a pterosaur can be built, it will be necessary to have as complete a picture of these animals as possible. Much of the conference was therefore devoted to

reviewing the most recent findings.

We now know that pterosaurs' wings were attached, not bat-like to the feet, but to the pelvis, to give a much more attenuate planform (Wellnhofer, P. *Abh. Bayer. Akad. Wiss.* **141**, 1; 1979); the resultant wing loading resembles closely that of modern fliers (Padian, K. *Discovery* **14**, 20; 1979). All pterosaurs, except the largest ones (like the largest birds), were active fliers; their wings were fully folded when they walked on the ground, which they did bipedally: their forelimbs are functionally homologous to those of their closest relatives, the dinosaurs and birds, not to those of the bats (Padian, K. *Paleobiology* **9**, 218; 1983).

Recent models have treated the wing membrane as if it billowed with no internal structure, and ignored the probability that most pterosaurs having a wingspan of less than 3 m seldom glided or soared, but flapped continually. Moreover, finely preserved specimens of *Rhamphorhynchus*, from the Late Jurassic of West Germany, show a series of fine, intercalated and stiffened striae permeating the entire wing membrane (Wellnhofer, *op. cit.*). This arrangement resembles the structural elements in the wings of birds (feather shafts) and bats (fingers), and suggests a role in maintaining the shape of the wing. The striae may have helped to determine the camber of the wing, although the details of their muscular control, including the extent to which the wing may have been collapsed without losing its aerodynamic integrity in flight, are not known.

Some idea must also be had of the dimensions and mobility of the animal to be modelled. This is complicated by the bimodal distribution of *Quetzalcoatlus* specimens. Thus there are around 20 partially-preserved 'small' individuals, with

wingspans of about 5.5 m. Taken collectively, they provide a complete three-dimensional picture of nearly every wing element (Langston, W. *Scient. Am.* **244**, 122; 1981). Then there is the giant specimen, known initially from its tree-like 50 cm long humerus. Each of its incomplete skeletal elements is approximately twice the length of its homologue in the 'small' form, although its shape is much more robust. Doubling the dimensions of the smaller form is not adequate for some parts of the body, including the torso, of which there are very few remains. Body length can be extrapolated from the few dorsal vertebrae preserved, but body weight is more difficult to determine and estimates range from 50 to 100 kg; until a scale replica of the animal can be built from preserved material, a weight of about 65 kg has been provisionally accepted. The wing area must depend on the length of the body, and seems to have been about 8.13 m²; this yields a wing loading of approximately 0.8 g cm⁻². *Quetzalcoatlus* had an aspect ratio of about 15 which, although high by the standards of natural fliers, is significantly lower than that of *Pteranodon* and other smaller seagoing pterosaurs. No living or man-made flier is an exact analogy for pterosaurs, although ultralight sailcraft may serve the purpose (McMasters, J. *NASA Conf. Pap.* **2085** (II); 1979).

The practical problems of replicating the flight of such an animal in a radio-controlled soarer-ornithopter are just beginning to be addressed. The first is that *Quetzalcoatlus*, like all pterodactyloid pterosaurs, had no tail, so pitch control will be complicated. Its neck and head were, by contrast, remarkably long, each about 2 m, while its entire torso was much shorter and its legs did not trail far behind the body. How far the neck could have been retracted or bent, if at all, is an important problem for paleontologists and has implications for engineers for calculating the centre of mass. The long neck and head form a powerful lateral moment arm, but also create a large amount of inherent instability; some sort of balancing mechanism up front will be necessary to control this. There are also logistic problems in trying to construct joints that can, for example, trim the wings and adjust pitch, while also supplying power for climbing flight.

The aim of the project is to launch the model from the Air and Space Museum of the Smithsonian Institution, from where it will fly across the Mall, climbing under its own power, circle the Washington Monument and return. To add realism, it has been suggested the model might pause in its flight, swoop down to snatch a small child from the crowd, carry it aloft, and consume it. For the present, mercifully, that objective seems to be beyond the technological capabilities of the field of Robotics.

Kevin Padian is Assistant Professor of Paleontology at the University of California, Berkeley, California 94720.

*9-10 July 1984, organized by Paul MacCready and sponsored by the National Air and Space Museum, Washington.

Geophysics

Continental interior in view

from Roger N. Anderson

ANOTHER frontier of geological exploration may at last be yielding to curiosity and advancing technology. After the great geological expeditions to the Moon, and dramatic discoveries of the ocean floor made over the last twenty years, the attention of a surprisingly large number of geologists worldwide is turning back to the original geological problem: the internal workings of the continents.

We know very little about how the continents form and deform. What we infer comes from remote-sensing with the use of classical geophysical and geochemical techniques such as seismology, gravity, petrology and mineralogy. But the sources of these observations are always at the surface. Although we can build elaborate models of the continental interiors, we shall never be able to verify these concepts without looking inside the continents. Experiments to this end are now being planned in several countries. The tool used to sample the inside of the Earth is the drillbit.

A remarkable coincidence of planning means that several countries will begin programmes of deep crustal drilling in 1986. German plans, outlined by E. Seibold at a conference* earlier this year, are to drill to at least 7 km through the Hycenian overthrust in the Black Forest. The aim is to understand how huge masses of rock could have slid several hundred kilometres over only slight slopes in the geological past. High pore pressures are suspected essentially to have floated these masses over other rock, but not a single measurement of pore pressure is yet available from such a shear zone. Scientists in Sweden, the United States, France, Austria and Japan are also planning to begin deep continental drilling in 1986.

Why are so many national programmes focusing on the drillbit at about the same time? Technology is probably the major driving factor. The international oil industry has developed in the drillbit a remarkable technology ideally suited for geological purposes as well as for hydrocarbon exploration. F. Schuh (Atlantic Richfield), for example, presented the conference with an entirely feasible design for a 50,000 foot well into crystalline rock. Several speakers (including G. Olhoeft and B. Dennis, US Geological Survey; P. Lysne, Sandia; M. Mathews, Los Alamos) offered a staggering array of geophysical and geochemical measurements and *in situ* experiments that can now be accomplished in a deep hole. It is now possible to measure

accurately the pore pressure, permeability, electrical and thermal conductivity, sonic, ultrasonic and seismic character of the wellbore as well as the exact chemical composition of the rock and pore fluids that are being drilled through.

It was not just these visions of technology that captured imaginations at the conference but also the accomplishments of continental drilling projects already well underway. Although not in attendance, the Russians dominated this phase of the meeting. They have been drilling deep into the continent for over 10 years, and at one hole on the Kola peninsula, they have made extraordinary discoveries. The current dogma holds that the deep crust is void of minerals and fluids because fractures, abundant at the surface, cannot remain open to fluid flow at great pressures. The Russians have found just the opposite: at 13 km depth, they encountered open fractures, mineralization and abundant evidence of fluid transport. They even discovered liquid sulphur, which is hydrolysed in the drilling mud to form sulphuric acid which eats away their drillbit.

The highlight of the conference was P. Robinson's (Dalhousie University) report of results from an international group, supported by Canada, the UK and US, that is drilling through the Troodos Ophiolite on Cyprus. Deep penetration of the pillow basalts and underlying sheeted dike com-

plex, has shown that Troodos is made neither from ordinary sea floor, nor from back-arc basin sea floor as has long been thought, but is instead the inside of an island arc. Moreover, M. Salisbury (Scripps Institution of Oceanography), using technology supplied by the British company British Plaster Board, has conducted a complete suite of geophysical measurements inside the hole, or logs. This revealed that the internal plumbing of an island arc produces layered structures almost identical to the crustal layers of the ocean floor. Hence the mistaken origin of the complex.

By August 1984, this collaboration had drilled to within 180 m of the MOHO, the seismic discontinuity between the Earth's crust and mantle. The penetration of that barrier — originally the aim of the Deep Sea Drilling Project — now looks likely to be achieved first by the Troodos drilling project. We eagerly await information on what lies beneath it.

The conference ended with an air of great optimism. Participants had heard from Bob Luth (Sandia) of a greater than 1 m penetration of magma inside Kiluea Volcano, Hawaii, and from W. Elders (University of California, Riverside) that drilling was about to begin in the 350°C waters of the Salton Sea geothermal area. Our understanding of the Earth was revolutionized by plate tectonics, which arose from exploration of the ocean floor. The exploration of the interior of the continents promises to further that understanding by a similar degree. □

Roger N. Anderson is in the Lamont-Doherty Geological Observatory of Columbia University, Palisades, New York 10964.

Archaeology

Kin-groups in Megalithic burials

from C.J. Scarre

MEGALITHIC tombs are among the most impressive neolithic monuments of western Europe, and a principal source of information about social organization in that period. Communal burial was the prevalent rite and in cases in which a tomb has not suffered too severely from the attentions of nineteenth-century antiquarians or been destroyed by a landowner, the archaeologist can expect to find large quantities of human remains. Despite their evident potential, however, analyses of such bone assemblages all too often fail to go beyond estimating the number and stature of the individuals represented and commenting on any obvious traces of injury and disease. How much more information this material holds is vividly demonstrated by the recent work at La Chaussée-Tirancourt, where the study of osteological abnormalities has allowed a series of distinct burial populations to be identified which may correspond to the different kin-groups using

the tomb.

La Chaussée-Tirancourt lies a few kilometres west of Amiens and is a megalithic monument of the 'allée couverte' type, with a burial chamber 9 m long and 2 m wide. By careful excavation over several years, Claude Masset and his team have recovered the remains of 356 individuals, 60 in layer V towards the bottom of the chamber and almost 300 in layer III higher up¹ (see figure). Within layer V, four separate 'zones' of burials were identified, while in layer III the human remains were concentrated in 'cases', which probably had originally been compartments or containers of wood or other organic material long since perished. At the end of its use the tomb was sealed by a substantial deposit of chalk and earth, on the surface of which fierce fires were lit, leaving a thick layer of ash. Curiously, the customary megalithic capstones are absent in this instance: the roofing of the chamber before the final

*Observation of the continental crust through drilling, supported by the US Department of Energy, Geological Survey and National Science Foundation, held in Tarrytown, New York, 20-25 May, 1984.

burning must instead have been of wood or thatch. Objects associated with the burials and radiocarbon dates from the layer of ash suggest that the monument was used actively between the late third and early second millennia BC.

Recently, Masset², Amaya Rodriguez Hernandorena³ and Pascal Sellier⁴ have carried out detailed analyses of the human material recovered from the tomb, concentrating on four bones (tibia, femur, humerus and calvarium) and on some fifty discrete (that is, non-metric) characters that are probably genetically determined. In both of the principal burial layers, the distribution of the characters among different parts of the tomb chamber was found to be significantly non-random (see figure).

In layer V, the remains from zones 2 and 4 form a single group characterized by a high incidence of incrustation of the humeral trochlea. Zone 1 is distinguished by the rarity of a third trochanter on the femur, a feature fairly common in the layer as a whole, and by an unusually high frequency of hypotrochanterian fossae. These and similar abnormalities allow the burials in layer V to be divided into three distinct populations: zones 1, 2 and 3/4.

Layer III shows a comparable, though somewhat more complex, patterning as the cases provide eight separate burial units. Case ω is distinguished by the frequency of an additional articular facet on the tibia. Cases ϵ , β and ζ share the absence of an olecranian perforation in females and the common presence of a hypotrochanterian fossa, but only case ϵ has a relatively high incidence of incrustation of the humeral trochlea.

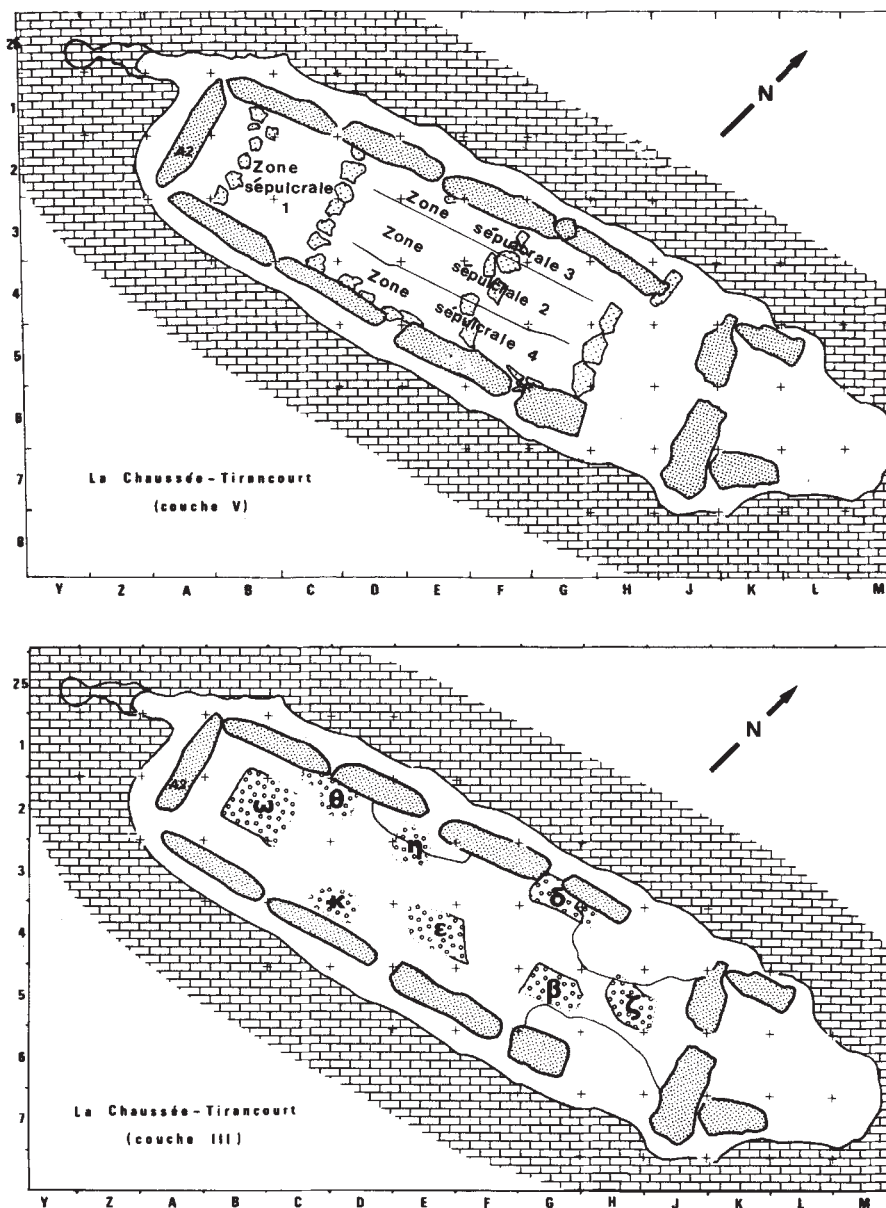
The presence of the same abnormalities in the bone assemblages of layers III and V suggest that the populations of the two layers were probably related, though some characters are more common in one than in the other, and the individuals in layer III are on the whole less robust and closer to the norm than those in layer V.

A recent study argues that megalithic tombs of the *allée couverte* type in north eastern France served as community foci

100 years ago

PROF. Ira Remsen testified to the utility of valence. He remarked that there were two ways of teaching: one by giving all the principal theories first, and the other giving the facts and then the theories — which latter he considered the best method. He had come to the conclusion that valence should never be mentioned until all the important properties of a compound are known. In regard to its value to young students, he thought its use was dangerous until they fully understood its meaning. He believed that the value of valence had been magnified, and that it was better to study the reactions of compounds, and the methods for their synthesis, and the manner of breaking-up.

Mr A.H. Allen said that many formulae that showed the structure of compounds according to the valence of the elements do not give any idea of the true constitution of these compounds as ascertained from their properties.



The megalithic tomb of La Chaussée-Tirancourt, showing the arrangement of the burials of the two principal layers in 'zones' (layer V, top) and in 'cases' (layer III, bottom). From ref. 2 and reproduced by permission of C. Masset.

for populations living in scattered hamlets or farmsteads round about, the burial of members of the different households in the central tomb operating as a crucial mechanism of social integration⁵. At La Chaussée-Tirancourt, the concentration of individual abnormalities in the different zones and cases suggest that each of these units was reserved for a separate kin or household group. This reservation of particular parts of the chamber for specific kin-groups may have persisted throughout the life of the tomb, since, for instance, incrustation of the trochlea is a common feature of humeri from the south-central sector both in layer III and in layer V. The spatial arrangement of the burials of related individuals in the tomb therefore reflects the scatter of family or household groups in their small settlements.

The research on the bone material from La Chaussée-Tirancourt clearly demon-

strates the potential use of genetically-dependent osteological abnormalities for improving our understanding of pre-historic burial practices. It is hoped that the excavation currently being conducted by Masset at Méréaucourt, another northern French *allée couverte*, will amplify the picture and throw further light on the social organization of the populations that used these monuments. □

1. Masset, C. *Antiquity* 46, 297 (1972).
2. Masset, C. *Arch. Suisses d'Anthropologie gen.* 43(2), 223 (1979).
3. Rodriguez-Hernandorena, A. *Etude des Humérus, Fémurs et Tibias de la Chaussée-Tirancourt (sépulture Seine-Oise-Marne de la Somme)* (diploma, Ecole Pratique des Hautes Etudes, 3ème section, Paris, 1983).
4. Sellier, P. *Les Caractères Non-métriques du Calvarium Humain. Application à l'étude d'une Population Archéologique* (thesis, Univ. Paris VI, 1983).
5. Howell, J.M. *Settlement and Economy in Neolithic Northern France* (British archaeol. Rep., Oxford, 1983).

C.J. Scarre can be contacted at 28 Priory Road, Cambridge CB5 8HT.

Parkinson's disease

Clues to aetiology from a toxin

from Solomon H. Snyder

UNTIL recently Parkinson's disease was one of few neurological conditions whose pathophysiology was presumed to be 'solved'. Degeneration in the corpus striatum, which regulates motor activity, leads to rigidity, difficulty in moving, and tremors of the limbs. Hornykiewicz, in 1960, showed that the neurotransmitter dopamine is depleted from the corpus striatum of affected patterns¹ and Cotzias later developed a therapy based on orally administered L-DOPA, the amino acid precursor of dopamine². At the same time, histochemical studies showed that the major dopamine pathway in the brain has cell bodies in the substantia nigra with axons ascending to terminate in the corpus striatum³. An animal model of Parkinson's disease soon followed: when injected into the vicinity of the substantia nigra, 6-hydroxydopamine, which is readily oxidized in cells to a toxic quinone, selectively accumulates through a pump-like mechanism exclusive to dopamine neurones with the result that the nigrostriatal dopamine pathway is destroyed.

This model, however, sheds no light on the aetiology of Parkinson's disease and may not even be adequate as a model of the human disease, since Parkinsonian brains show damage in areas other than the substantia nigra, including the noradrenaline-containing locus coeruleus. Several recent papers, two of them just published in *Nature*^{4,5} show how a new animal model, based on the toxicity of 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), may overcome both shortcomings.

The toxin was originally implicated through a bizarre detective story. In the late 1970s a chemistry graduate student suddenly developed severe Parkinsonian symptoms, which were peculiar because of the patient's youth and the abrupt onset of the disease. The diagnosis was confirmed, following the patient's death from a drug overdose, when an autopsy revealed selective degeneration of the substantia nigra. Astute research scientists at the National Institutes of Mental Health in Bethesda learned that the patient had been synthesizing, for illicit sale, a close relative of the well-known opiate, meperidine. Chemical analysis of samples recovered from the patient's basement laboratory revealed contaminants, especially MPTP⁶. Subsequently, Californian investigators identified several patients with virtually identical clinical symptoms who had also taken preparations containing MPTP⁷.

These cases have prompted speculation that MPTP or similar environmental toxins might 'cause' most cases of Parkinson's disease⁸. Direct ingestion may not be necessary: recently a 37-year-old chemist

developed serious irreversible Parkinson's disease while using MPTP as a synthetic intermediate⁹, but without ever ingesting it. Inhalation or skin contact may have sufficed to elicit the condition. Since MPTP itself is not toxic to cells, it is presumably converted to a more chemically reactive form. This fits the well-documented observation that cigarette smokers are less prone to Parkinson's disease than non-smokers^{10,11}, in the sense that carbon monoxide, which accumulates in the tissues of smokers, may detoxify active intermediates produced from MPTP and related agents. If Parkinson's disease is due to environmental toxins, one need not expect the sudden onset that follows MPTP ingestion, nor would chronic, life-long exposure be required. Instead, initial damage to the substantia nigra, which might occur early in life, combined with the natural decrease with age of the dopamine-containing cells in the substantia nigra, could suffice to produce symptomatic Parkinsonism.

Several workers have used MPTP to develop an animal model of Parkinson's disease and have found species differences that have provided important clues to the mechanisms of toxicity. Unlike rats and guinea pigs¹²⁻¹⁵, rhesus monkeys¹² and squirrel monkeys¹⁶ given MPTP develop clinical Parkinsonian symptoms and show selective destruction of the substantia nigra. MPTP also produces irreversible depletion of striatal dopamine in mice¹⁷⁻¹⁹.

MPTP can be oxidized by mitochondrial enzymes such as monoamine oxidase and possibly aldehyde dehydrogenase to 1-methyl-4-phenylpyridine (MPP⁺), a charged pyridinium derivative²⁰, which is chemically reactive and toxic to tissues. It is detectable in the brain for prolonged periods after *in vivo* administration of MPTP⁴. An impressive link between the oxidation and neurotoxicity of MPTP has been reported by Heikkilä and associates⁵ who show that its neurotoxicity in mice can be prevented by pretreatment of the animals with monoamine oxidase-inhibiting drugs, which also block the conversion of MPTP to MPP⁺.

But how does one explain the selective effect of MPTP on brain structures such as the substantia nigra and corpus striatum? Although MPP⁺ is formed to a similar extent in all areas of the brain¹², species variations in recently identified receptor-like binding sites for MPTP may help provide an answer. MPTP binds with high affinity ($K_D = 2 \times 10^{-8} M$) to receptors in brains of numerous species. By autoradiographic techniques one can visualize MPTP receptors. In human brain they are most highly concentrated in the substantia

nigra and corpus striatum²¹ but we²¹ and others²² find substantially fewer receptors in the same regions of rat brain. Species variations in nigrostriatal receptor density thus may account for the lesser toxicity of MPTP in rats. Both human and rat brain have MPTP receptors in several other areas. One site of particularly high receptor density is the locus coeruleus. This localization fits well with the marked changes in noradrenaline metabolism that follow MPTP treatment of rats^{13,14} and monkeys¹², with the pronounced changes in accumulation of 2-deoxyglucose by the locus coeruleus in MPTP-treated rat and guinea pig²³ and with the degeneration of the locus coeruleus in Parkinson's disease.

The most recent revelation in the MPTP story comes from Parsons and Thomas C. Rainbow (a leader in autoradiographic studies of neurotransmitter receptors, who met an accidental death on 6 September 1984). They compared by autoradiography the localizations of binding sites for ³H-MPTP and for the monoamine oxidase inhibitor ³H-pargyline²⁴, which selectively labels the type B isozyme of monoamine oxidase. Localizations of the two binding sites were virtually identical, indicating that the MPTP receptor may be the type B enzyme monoamine oxidase itself.

These findings not only reinforce the role of monoamine oxidase in MPTP actions, but also raise the intriguing possibility that the selective receptor binding sites that have been reported for diverse drugs might be enzymes rather than receptors for endogenous neurotransmitters. The very discrete localizations observed for monoamine oxidase suggest a more selective role for this enzyme in brain function than was previously suggested. □

1. Hornykiewicz, O. *Pharmac. Rev.* **18**, 925 (1966).
2. Cotzias, C.G., Van Woert, M.H. & Schiffer, L.M. *New Engl. med. J.* **276**, 374 (1967).
3. Ungerstedt, U. *Acta physiol. scand. Suppl.* **367**, 1 (1971).
4. Markey, S.P., Johannessen, J.N., Chueh, C.C., Burns, R.S. & Herkenham, M.A. *Nature* **311**, 464 (1984).
5. Heikkilä, R.E., Manzino, L., Cabbat, F.S. & Duvoisin, R.C. *Nature* **311**, 467 (1984).
6. Davis, G.C. et al. *Psychiat. Res.* **1**, 249 (1979).
7. Langston, J.W., Ballard, P., Tetrud, J.W. & Irwin, I. *Science* **219**, 979 (1983).
8. Calne, D.B. & Langston, J.W. *Lancet* **ii**, 1457 (1983).
9. Langston, J.W. & Ballard, P.A. *New Engl. med. J.* **309**, 310 (1983).
10. Kessler, I.L. *Adv. Neurol.* **19**, 355 (1978).
11. Baumann, R.J. et al. *Neurology* **30**, 839 (1980).
12. Burns, R.S. et al. *Proc. natn. Acad. Sci.* **80**, 4546 (1983).
13. Chueh, C.C. et al. *Eur. J. Pharmac.* **100**, 189 (1984).
14. Enz, A., Hefti, F. & Frick, W. *Eur. J. Pharmac.* (in press).
15. Steranka, L.R., Polite, L.N., Perry, K.W. & Fuller, R.W. *Res. Commun. Substance Abuse* **4**, 315 (1983).
16. Langston, J.W., Forno, L.S., Rebert, C.S. & Irwin, I. *Brain Res.* **292**, 390 (1984).
17. Hallman, H., Olson, L. & Jonsson, G. *Eur. J. Pharmac.* **97**, 133 (1984).
18. Wallace, R.A., et al. *Life Sci.* **35**, 285 (1984).
19. Heikkilä, R.E., Hess, A. & Duvoisin, R.C. *Science* **224**, 1451 (1984).
20. Chiba, K., Trevor, A. & Castagnoli, Jr., N. *Biochem. biophys. Res. Commun.* **120**, 574 (1984).
21. Javitch, J.A., Uhl, G.R. & Snyder, S.H. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4591 (1984).
22. Wiecek, C.M., Parsons, B. & Rainbow, T.C. *Eur. J. Pharmac.* **98**, 453 (1984).
23. Palacios, J. & Wiederhold, K. *Brain Res.* **301**, 187 (1984).
24. Parsons, B. & Rainbow, T. *Eur. J. Pharmac.* **102**, 375 (1984).

Solomon H. Snyder is in the Departments of Neuroscience, Pharmacology and Experimental Therapeutics, Johns Hopkins School of Medicine, Baltimore, Maryland 21205.

Similarity between p15E of murine and feline leukaemia viruses and p21 of HTLV

SIR — Retroviral infections of a number of animal species are frequently associated with immunosuppression. The hydrophobic retroviral transmembrane protein p15E is highly conserved among type C and type D retroviruses¹. This protein may contribute to the immunosuppression associated with retroviral infections since it inhibits *in vitro* lymphocyte transformation in man², cats³ and mice⁴ and increases tumour incidence in cats given feline sarcoma virus⁵. The mechanism for the inhibition of human lymphocyte transformation by p15E was shown to be due to suppression of interleukin-2 (IL-2) production⁶. We have demonstrated *in vitro* suppression of human monocyte function⁶ and *in vivo* inhibition of murine macrophage accumulation to inflammatory foci by p15E from murine leukaemia virus (MuLV)⁷ or by tumour-associated proteins which are antigenically related to p15E^{6,8}.

Human T-cell leukaemia-lymphoma virus (HTLV) is the designation for a family of T-lymphotropic, exogenous type C retroviruses which have been isolated in a number of countries⁹⁻¹³. These viruses have been linked as aetiological agents to certain types of leukaemias and lymphomas and to the acquired immunodeficiency syndrome (AIDS)¹⁴⁻¹⁶. It has recently been reported that the polymerase gene products of HTLV, MuLV and avian Rous sarcoma virus are partially homologous in their amino acid sequence¹⁷. Since immunosuppression often accompanies the HTLV-associated T-cell malignancies and is characteristic of AIDS, we hypothesized that HTLV-viruses might have, in addition to their lymphocytopathic effects, other features in common with immunosuppressive MuLV and feline leukaemia viruses (FeLV). We therefore compared the amino acid sequences for p15E of Friend¹⁸, Moloney¹⁹, AKV²⁰, and Gross²¹ murine leukaemia viruses (FLV, MoLV, AKV, GLV), FeLV²², and mink-cell focus-forming viruses of Moloney²³ and AKR²⁴ origin (MMCF, AMCF) with the sequences of the envelope proteins of HTLV-I²⁵ and HTLV-II²⁶. The sequences were analysed for homology using the PROTHOM computer program developed by Fristensky *et al.*²⁷.

As shown in the figure above there is a significant degree of homology (73%) between the various p15E proteins and HTLV-I and HTLV-II which occurs in a 26-amino acid sequence located in the p21 region of HTLV. It is particularly noteworthy that the first 10 amino acids in this region are identical and that this region of homology occurs in almost the same region of the p15E and p21 molecules. The homology begins at residue 70 in p15E and at residue 377 of the HTLV-I envelope protein and at residue 373 of the HTLV-II envelope protein which both correspond to

HTLV _I ENV	Q N R R G L D L L F	W E Q	G G L C K	A L Q	E Q C R F	377	402
HTLV _{II} ENV	Q N R R G L D L L F	W E Q	G G L C K	A I Q	E Q C C F	373	398
FLV p15E	Q N R R G L D L L F	L K E	G G L C A	A L K	E E C C F	70	95
MLV p15E	Q N R R G L D L L F	L K E	G G L C A	A L K	E E C C F		
AKV p15E	Q N R R G L D L L F	L K E	G G L C A	A L K	E E C C F		
GLV p15E	Q N R R G L D L L F	L K E	G G L C A	A L K	E E C C F		
MMCF p15E	Q N R R G L D L L F	L K E	G G L C A	A L K	E E C C F		
AMCF p15E	Q N R R G L D L L F	L K E	G G L C A	A L K	E E C C F		
FeLV p15E	Q N R R G L D L L F	L K E	G G L C A	A L K	E E C C F		

Amino acid sequence homology between p21 of HTLV-I and HTLV-II and p15E of murine and feline leukaemia viruses. Residue 377 of the HTLV-I envelope and 373 of the HTLV-II envelope both correspond to residue 65 of the p21 protein.

residue 65 of the p21 transmembrane protein²⁶.

The significance of such a well conserved region occurring in murine, feline, and human retroviruses is not yet clear. However, since dramatic immunosuppression is often associated with these viruses and purified p15E is immunosuppressive in a variety of systems, it will be particularly important to determine whether p21 of HTLV can also inhibit immune functions.

GEORGE J. CIANCIOLO

ROBERT J. KIPNIS

RALPH SNYDERMAN

Howard Hughes Medical Institute,
Department of Medicine,
Duke University Medical Center,
Durham,
North Carolina 27710, USA

- Thiel, H.-J., Broughton, E.M., Matthews, T.J., Schafer, W. & Bolognesi, D.P. *Virology* **111**, 270-274 (1981).
- Copelan, E.A. *et al. J. Immun.* **131**, 2017-2020 (1983).
- Mathes, L., Olsen, R.G., Hebebrand, L.C., Hoover, E.A. & Schaller, J.P. *Nature* **274**, 687-689 (1978).
- Fowler, A.K., Twardzik, D.R., Reed, C.D., Weislow, O.S. & Hellman, A. *Cancer Res.* **37**, 4529-4531 (1977).
- Mathes, L.E. *et al. Cancer Res.* **39**, 950-955 (1979).
- Cianciolo, G.J., Hunter, J., Silva, J., Haskill, J.S. & Snyderman, R. *J. clin. Invest.* **68**, 831-844 (1981).
- Cianciolo, G.J., Matthews, T.J., Bolognesi, D.P. & Snyderman, R. *J. Immun.* **124**, 2900-2905 (1980).
- Cianciolo, G.J., Lofstrom, M.E., Tam, M. & Snyderman, R. *J. exp. Med.* **158**, 885-900 (1983).
- Poiesz, B.J., Russetti, F.W., Gazdar, A.F., Bunn, P.A., Minna, J.D. & Gallo, R.C. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7415-7419 (1980).
- Gallo, R.C. *et al. Cancer Res.* **43**, 3892-3898 (1983).
- Haynes, B.F. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 2054-2058 (1983).
- Miyoshi, I. *et al. Nature* **296**, 770-771 (1981).
- Popovic, M. *et al. Science* **219**, 856-859 (1983).
- Gallo, R.C. *et al. Science* **224**, 501-503 (1984).
- Popovic, M., Sarnagadharan, M.G., Read, E. & Gallo, R.C. *Science* **224**, 497-500 (1984).
- Sarnagadharan, M.G., Popovic, M., Bruch, L., Schupbach, J. & Gallo, R.C. *Science* **224**, 506-508 (1984).
- Patarca, R. & Haseltine, W.A. *Nature* **309**, 288 (1984).
- Koch, W., Hunsmann, G. & Friedrich, R. *J. Virol.* **45**, 1-9 (1983).
- Shinnick, T.M., Lerner, R.A. & Sutcliffe, J.G. *Nature* **293**, 543-548 (1981).
- Lenz, J., Crowther, R., Straceski, A. & Haseltine, W. *J. Virol.* **423**, 519-529 (1982).
- DesGroselliers, L., Villemur, R. & Jolicoeur, P. *J. Virol.* **47**, 24-32 (1983).
- Elder, J.H. & Mullins, J.I. *J. Virol.* **46**, 871-880 (1983).
- Bosselman, R.A., van Straaten, F., Van Beveren, C., Verma, I.M. & Vogt, M. *J. Virol.* **44**, 19-31 (1982).
- Kelley, M. *et al. J. Virol.* **45**, 291-298 (1983).
- Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3618-3622 (1983).
- Sodroski, J. *et al. Science* **225**, 421-424 (1984).
- Fristensky, B., Lis, J. & Wu, R. *Nucleic Acids Res.* **10**, 6451-6463 (1982).

Islet activating protein analogous to p21 ras?

SIR — In a recent article on mutant *ras* protein and cell transformation, Rob Newbold considered the significance of new findings about p21 *ras* function and activation. In particular, similarities between p21 *ras* and the guanine nucleotide binding proteins which are responsible for stimulation and inhibition of adenylate cyclase were emphasized and an analogy drawn with the action of cholera toxin on G_s protein GTPase, implying that p21 *ras* might act in an analogous manner to cholera toxin, which exerts a strong proliferative stimulus on certain mammalian cells in culture.

I suggest that the analogy with cholera toxin is dubious in view of the fact that rat cells transformed by Kirsten sarcoma virus have been shown to contain less cyclic AMP and lowered adenylate cyclase activity. Therefore a closer analogy may be with the action of islet activating protein which reacts with Gi proteins to inhibit adenylate cyclase and reduce cyclic AMP. This would imply that decreased cyclic AMP assists proliferation.

R. K. RALPH

Department of Cell Biology,
The University of Auckland,
Private Bag, Auckland, New Zealand

- Newbold, R. *Nature* **310**, 628-629 (1984).
- Carchman, R.A., Johnson, G.S., Pastan, I. & Scolnick, E. *M. Cell* **1**, 59-64 (1974).
- Anderson, W. B., Gallo, M. Pastan, I. *J. biol. Chem.* **249**, 7041-7048 (1974).
- Gilman, A. *Cell* **36**, 1577-1579 (1984).

Heritable IQ — a reason to bother

SIR — The recent article from Teasdale and Owen¹ prompts one to wonder about IQ genes. If intelligence is substantially heritable, is it determined by genes in a positive sense, such that better allelic combinations of IQ genes produce more intelligent people? This may seem to be the most parsimonious explanation, but parsimony is in the eye of the beholder.

That simple genetic mechanisms can account for conspicuously low IQ stirs no

controversy (such as, phenylketonuria, Down's syndrome, Tay-Sachs), and most would agree that relatively mild homozygotic conditions can depress the IQ of afflicted persons (children who suffer from sickle cell anaemia or cystic fibrosis may test lower for IQ, on average, simply because they miss a great deal of school). Does this mean that such children have genetically heritable IQ? Does it mean, in the formal sense, that the disparity in IQ between them and normal children is attributable to variance in genes? Of course, it means exactly that — and yet the genes for beta haemoglobin and regulation of mucous production are not "IQ genes".

I am certain that my results on an IQ test would suffer if I were tested while I had a head cold. Imagine two boys reared in the same environment who are genetically identical except that one has an inherited sensitivity to ragweed. If they lived in a high ragweed area, such that the allergic one suffered from chronic congestion, would it be surprising to find that he tests slightly lower for IQ? Intelligence is a summary performance indicator, as distinct from an isolated component indicator (like how well a person's brain works), in the sense that suboptimal performance in any subsystem can have a negative impact. Variability in intelligence is perhaps analogous to the variability in horse power of a group of nominally identical engines, where quality of fuel represents the environmental component and fine tuning represents the genetic component. Performance is maximized when an engine is tuned for its operating conditions. A set of engines tuned in the tropics will not perform as well as when transported to Toronto. The question becomes, to what degree does the distribution of non-optimal non-IQ-genes account for heritable IQ?

The relative optimality of a gene is often environmentally specified. Teasdale and Owen, like others before them, found that the heritable component of IQ could be determined by segregation of a few alleles. ("The results for intelligence conform closely to what would be predicted by a simple polygenic model of genetic transmission . . ."). There are two ways this could happen: (1) there may be a universal handful of IQ genes that segregate appropriately, or (2) in any given environment, there may be a handful of genes for each of any one of hundreds of traits that can have a negative impact on IQ, and some of those handfuls segregate appropriately. If the former is the case, one wonders why alleles for superior IQ are not closer to fixation. Is there polymorphism for IQ because there were evolutionary environments in which intelligence was selected against?

Think about breathing through your nose in a steam bath. This is a genuine problem. Indeed, if you have a relatively long, thin, hairy nose (as typifies people

from temperate zones and/or high altitudes — like Swedes and Ethiopians) it can be virtually impossible to pull enough air through that nose. But observe the steam bath breathing behaviour of someone with a wide, short, hairless nose (as typifies blacks and orientals indigenous to tropical wetlands). These people can breathe superhumid air through their nose with no problem. Under the extreme condition of a steam bath, nose shape so strongly affects breathing efficiency that you can feel the difference by mechanically flaring your nostrils. Humans take over 10 million breaths a year. I submit that if we were all reared in steam baths, people with wide, short, hairless noses would have slightly higher mean IQ. This scenario is not silly because real noses in the real world are very differently adapted to conditions of humidity or dust. Among American racial and ethnic groups, only blacks and non-Japanese orientals have a conspicuous prevalence of lung disease³. Dust and airborne microorganisms can be a serious health hazard⁴. Noses have evolved accordingly.⁵⁻⁷

Can differences in nose shape (inherited) account for differences between black and white American IQ? It would be ridiculous to think so, but it should be considered equally ridiculous to assume that such factors have no effect. What about skin colour? More sunlight is required by dark skin to synthesize the same amount of vitamin D. Just as dark skin prevents hypervitaminosis-D in the tropics, so light skin facilitates D synthesis in temperate zones⁸⁻¹⁰. The question becomes, how

many such factors are there? Dozens? Hundreds? More than enough to account for the portion of IQ variance which is heritable? Quite possibly. But this is not why the question is important.

The question is important because there are many factors that something could be done about if we knew what they were. People with dark skin in dark climates can get some extra sunshine. People with non-dust-trapping noses can have humidifiers in their homes during winter and take care to wear a dust mask when doing a dusty job. Unfortunately, looking for genetic differences that incidentally affect IQ is politically unacceptable because such investigations tacitly acknowledge the existence of such differences.

It may seem naive to think that things like clean air and sunshine could ameliorate genetic differences that contribute to differences in IQ. Fortunately, naivete is also in the eye of the beholder.

JOHN HARTUNG

Department of Anesthesiology,
Downstate Medical Center,
State University of New York,
New York 11203, USA

1. Teasdale, T.W. & Owen, D.R. *Nature* **309**, 620-622 (1984).
2. Hartung, J. *Curr. Anthropol* **21**, 131-132 (1980).
3. Damon, A. *Am. J. Phys. Anthropol.* **22**, 375-382.
4. Cohen, D., Arai, S.F. & Brain, J.D. *Science* **204**, 514-517 (1979).
5. Newman, M.T. *Am. Anthropol.* **55**, 311-327 (1953).
6. Biasutti, R. *Razze e Popoli della Terra* (Unione Tipografico-Edi Traice, Torino, 1959).
7. Wolpoff, M.H. *Am. J. Phys. Anthropol.* **29**, 405-423 (1968).
8. Loomis, W.F. *Science* **157**, 501-506 (1967).
9. Haddad, J.G. & Hahn T.J. *Nature* **244**, 515-516 (1973).
10. Robertson, I., Glekin, B.M., McIntosh, W.B., Lakhani, A. & Dunnigan, M.G. *Proc. Nutr. Soc.* **41**, 243-250 (1982).

Raising the tone of DNA sequences

SIR — We assessed the potential for gene jingles following a lunchtime conversation two years ago. Hayashi and Munkata (*Nature* **310**, 96; 1984) appear to have used a purely arbitrary pitch-to-base assignment and ignored the complementary DNA strand altogether. We chose the notes D, A, F sharp, D, but have given some thought to base assignment of which there are 24 possibilities. We felt the complementary base triplet of each codon should produce a 'complementary' tune. Rather naively, we suspect, in musical terms, we took as complementary the quasi-mirror image of the notes on the stave (on the left of the figure). Eight of the possible 24 pitch-to-base assignments meet this requirement, but the assignment used by Hayashi and Munkata does not (on the right).

One of us (M.P.) sent short recorded DNA tunes to the other for the final selection of the pitch-to-base assignment by ear. No one assignment stood out as best, but the tunes of the promoter region CCAAT box were more pleasing musically than the 9-base repeat of collagen.

We are working towards the first public performance of 'Variations on a Cryptic Splice Junction' and the correspondence has given us renewed enthusiasm.

MARCUS PEMBREY

Mothercare Unit of Paediatric Genetics,
Institute of Child Health,
30 Guilford Street,
London WC1 1GH

DAVID HART

8 Chester Street,
London SW1, UK

Base pairing by Pembrey and Hart (left) and Hayashi and Munkata (right).

Formation of galaxies and large-scale structure with cold dark matter

George R. Blumenthal* & S. M. Faber*

* Lick Observatory, Board of Studies in Astronomy and Astrophysics, University of California, Santa Cruz, California 95064, USA

Joel R. Primack†§ & Martin J. Rees‡§

† Stanford Linear Accelerator Center, Stanford University, Stanford, California 94305, USA

‡ Institute of Theoretical Physics, University of California, Santa Barbara, California 93106, USA

The dark matter that appears to be gravitationally dominant on all scales larger than galactic cores may consist of axions, stable photinos, or other collisionless particles whose velocity dispersion in the early Universe is so small that fluctuations of galactic size or larger are not damped by free streaming. An attractive feature of this cold dark matter hypothesis is its considerable predictive power: the post-recombination fluctuation spectrum is calculable, and it in turn governs the formation of galaxies and clusters. Good agreement with the data is obtained for a Zeldovich ($|\delta_k|^2 \propto k$) spectrum of primordial fluctuations.

WHY are there galaxies, and why do they have the sizes and shapes that we observe? Why are galaxies clustered hierarchically in clusters and superclusters, separated by enormous voids in which bright galaxies are almost entirely absent? And what is the nature of the invisible mass, or dark matter, that we detect gravitationally roundabout galaxies and clusters but cannot see directly in any wavelength of electromagnetic radiation? Of the great mysteries of modern cosmology, these three may now be among the ripest for solution.

Because there is evidence that the mass of dark matter in the Universe exceeds that of the visible matter by at least an order of magnitude¹, the third question may hold the key to the first two. We now consider the hypothesis that the dark matter is cold²⁻⁷, that is, that its thermal velocity is cosmologically negligible in the early Universe.

Most modern theories of the origin of structure in the Universe assume that the extreme inhomogeneity of the present Universe grew gravitationally from initially very small density fluctuations^{8,9}. One hypothesis extensively explored¹⁰⁻¹² is that the dark matter (DM) consists of neutrinos of mass ~ 30 eV. Because these particles are noninteracting for $t \geq 1$ s but are still relativistic at $t \sim 1$ yr when galaxy-size masses first come within the horizon ($R_{\text{horizon}} = ct$), they would freely stream away, smoothing out fluctuations of horizon size or smaller. We refer to such particles as hot DM, and to the phenomenon just mentioned as damping by free streaming. The mass of the neutrinos inside the horizon when they first become nonrelativistic is roughly $10^{15} M_{\odot}$, about the mass of a supercluster ($M_{\odot}$ is the solar mass). This is consequently the mass of the first structures to collapse gravitationally in a neutrino-dominated Universe.

It is also possible that the DM consists of elementary particles that decoupled thermally from the big bang much earlier than neutrinos. It can be shown that such particles would have lower number density today, and thus could be more massive (~ 1 keV) and become nonrelativistic sooner than neutrinos^{13,14}. We refer to this as warm DM. The first structures to form in such a Universe would weigh about $10^{11} M_{\odot}$, about the mass of a typical galaxy¹⁵⁻¹⁷.

In contrast, in a Universe dominated by cold DM, the free streaming damping mass is, by definition, smaller than galaxy masses. Below, we will discuss possible particle physics candidates for cold DM and explain why we believe it is more

plausible that the Universe is dominated gravitationally by cold DM than by ordinary matter (baryons) or hot or warm DM. We next consider galaxy formation in the cold DM picture, and show that galaxy and cluster data compare favourably with the predictions of this model. We then discuss the extent to which the data constrain the cosmological density in the cold DM picture, and the evolution of superclusters and voids. The cold DM hypothesis seems to lead to a remarkably attractive theory for galaxy formation and to account for large-scale structure at least as well as any competing theory.

Evidence that dark matter is cold

Cold DM consists of particles having negligible thermal velocity with respect to the Hubble flow and having nongravitational interactions that are much weaker than the weak interactions^{3,4}. A popular cold DM candidate is the axion¹⁸, a pseudoscalar field proposed originally to avoid large CP (constant parity) violation in the strong interactions (which would imply, for example, much too large a value for the neutron electric dipole moment)¹⁹⁻²¹. Instanton effects generate a nonzero axion mass at the quark deconfinement temperature ($T \sim 10^2$ MeV), below which the axions act as a nonrelativistic, massive, pressureless fluid. The requirement that the axion density be less than the critical density implies that the axion mass $m_a \geq 10^{-5}$ eV (refs 22-24), while the longevity of helium burning stars implies that $m_a < 10^{-1}$ eV (ref. 25). Thus, if axions exist, they may be cosmologically important, and, for $m_a \approx 10^{-5}$ eV, they would be gravitationally dominant. If axions comprise the dark halo of our Galaxy, laboratory experiments have recently been proposed that could detect them²⁶.

Another cold DM candidate particle is the photino, the spin $\frac{1}{2}$ supersymmetric partner of the photon. Photinos are thought to be the lightest supersymmetric particle, with $m_{\tilde{\gamma}} \geq 0.5$ GeV (the lower limit corresponding to cosmological critical density)²⁷. As photino annihilation at high temperatures is incomplete, the remnant photinos can, because of their large mass, contribute a critical density today.

A third cold DM candidate is black holes^{28,29} of mass $10^{-16} M_{\odot} \leq M_{\text{BH}} \leq 10^6 M_{\odot}$, the lower limit implied by the non-observation of γ rays from black hole decay by Hawking radiation, and the upper limit required to avoid disruption of galactic disks and star clusters^{30,31}. Stronger but more controversial upper limits are $M_{\text{BH}} < 10^2 M_{\odot}$ from dwarf spheroidal haloes³² and $M_{\text{BH}} < 10^{-2} M_{\odot}$ from the non-observation of focusing of quasar cores³³.

§ Permanent addresses: Santa Cruz Institute of Particle Physics, University of California, Santa Cruz, California 95064, USA (J.R.P.); Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK (M.J.R.).

Another exotic cold DM candidate, recently proposed by Witten, is "nuggets" of u - s - d symmetric quark matter³⁴. There is thus no shortage of cold DM candidate particles—although there is no direct evidence that any of them actually exists. Our motivation for considering the hypothesis that the Universe is dominated by cold DM is twofold. First, there is more than a little evidence against baryonic, hot, and warm DM. Moreover, a cold DM Universe correctly predicts many of the observed properties of galaxies, including their range of masses, irrespective of the identity of the cold particle.

The most conventional assumption might be that DM is baryonic, since we know that baryons pervade the Universe. There are, however, three arguments that make a nonbaryonic form of DM more plausible. First, if DM consists of baryons, it must be clumped into objects with mass $\ll M_\odot$ to avoid nuclear burning and the emission of too much radiation³⁵, and there is no compelling theory for the formation of such a large density of objects of planetary mass or smaller. The difficulties with other possible forms of baryonic dark matter have been reviewed elsewhere³⁶. The second argument involves the observed deuterium abundance, $D/H = (1-4) \times 10^{-5}$ (by number), which provides a lower limit on the primordial D abundance because deuterium is readily consumed but not produced in stars. This then corresponds to an upper limit for the baryonic density parameter (ratio of baryon density to critical density) of $\Omega_b \leq 0.035h^{-2} (T_0/2.7)^3$ (ref. 37), where T_0 is the present temperature of the microwave background radiation and h is Hubble's constant expressed in units of $100 \text{ km s}^{-1} \text{ Mpc}$. (Observationally, $1/2 \leq h \leq 1$.) However, there is also strong observational evidence that the total density parameter $\Omega \geq 0.1$. Therefore a baryon dominated Universe ($\Omega = \Omega_b$) is consistent with the deuterium limit only for Ω and h very near their observational lower limits. Finally, the existence of galaxies and clusters today requires that perturbations in the density must have become nonlinear before the present epoch. In a baryonic Universe, for adiabatic perturbations at recombination, this implies present-day fluctuations in the microwave background much larger than the present observational upper limits^{38,39}. For baryonic DM, this problem can be avoided if the primordial fluctuation spectrum is isothermal rather than adiabatic, contrary to grand unified models of baryosynthesis^{40,41}, or if there is significant reheating of the intergalactic medium after recombination⁴². For nonbaryonic DM, on the other hand, the predicted fluctuations in the microwave background can be consistent with the observations as the fluctuations in the baryon-photon fluid are small at recombination and only later grow to the same size as fluctuations in the DM.

The neutrino dominated picture of galaxy formation also has serious weaknesses. Studies of nonlinear clustering (on scales $\lambda < 10 \text{ Mpc}$) show that supercluster collapse must have occurred quite recently, at redshift $z_{sc} < 2$ (refs 43, 44). This is also consistent with a study of streaming velocity in the linear regime ($\lambda > 10 \text{ Mpc}$), which indicates $z_{sc} < 0.5$ (ref. 45). However, the best limits on galaxy ages coming from globular clusters and other stellar populations, plus the possible association of QSOs with galactic nuclei, indicate that galaxy formation took place before $z = 3$ (ref. 46). This is inconsistent with the neutrino 'top-down' theory, in which superclusters form before galaxies rather than after them.

Another problem with the neutrino picture is that large clusters of galaxies can accrete neutrinos more efficiently than ordinary galactic haloes, which have lower escape velocities. One-dimensional numerical simulations predict that the ratio of total-to-baryonic mass M/M_b should be ~ 5 times larger for clusters ($\sim 10^{14} M_\odot$) than for ordinary galaxies ($M \sim 10^{12} M_\odot$)⁴⁷. While there is evidence that mass-to-light ratio M/L does increase with scale, there is also considerable evidence that the more physically meaningful ratio of total-to-luminous mass M/M_{lum} remains constant from large clusters through groups of galaxies, binary galaxies, and ordinary spirals. (M_{lum} , which is the mass visible in galactic stars and gas plus hot, X-ray emitting gas, is $\leq M_b$, as an unknown fraction of the baryons is

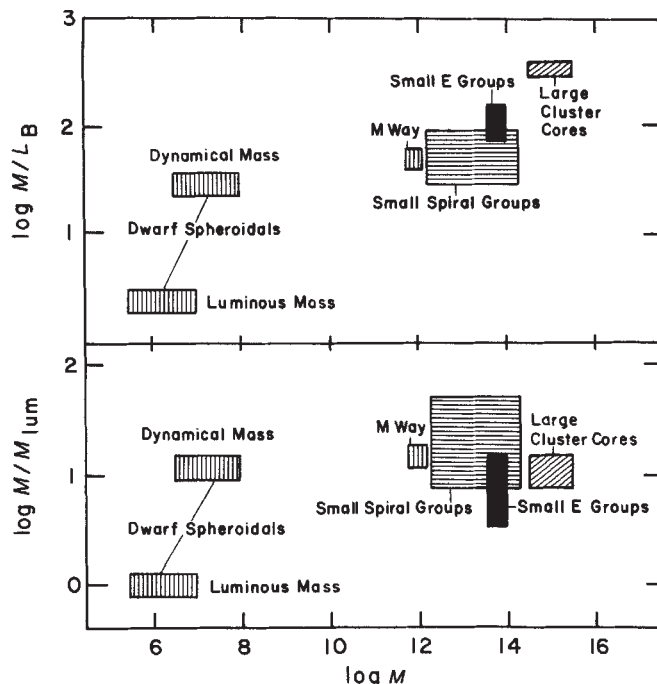


Fig. 1 Mass-to-light ratio, M/L_b , and total-to-luminous mass, M/M_{lum} , for structures of various size in the Universe. The data come from Table 1. Although M/L_b increases systematically with mass, the more physically meaningful ratio M/M_{lum} appears to be constant on all scales within the errors. If the velocity dispersion data for the dwarf spheroidal galaxies are interpreted to imply heavy haloes, the upper estimates result. The lower estimates follow from assuming that all the mass is visible. We believe the former estimate to be more realistic.

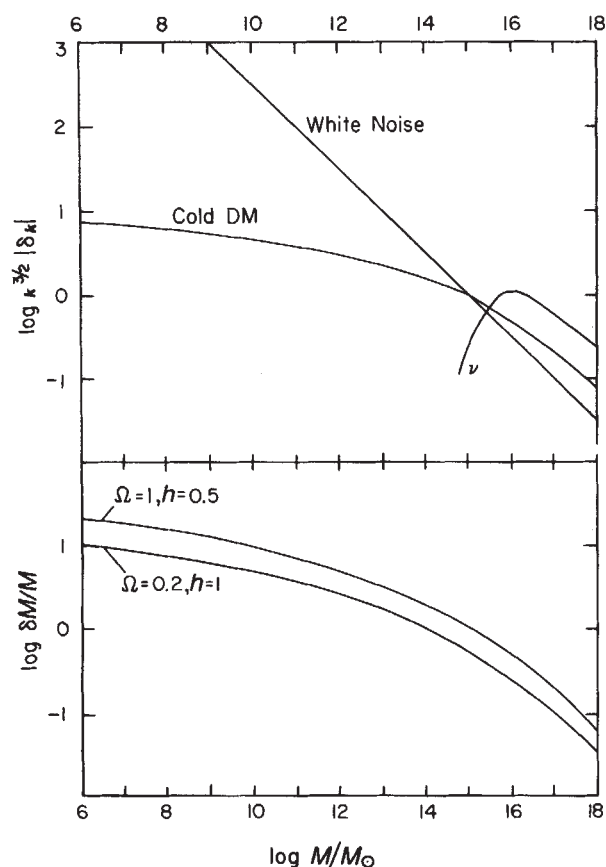


Fig. 2 Density fluctuations as a function of mass. a , $k^{3/2} |\delta_k| = \delta \rho / \rho(M)$, where $M = 4\pi^2 \rho_0 / 3k^3$, for isothermal white noise ($n=0$), and adiabatic Zeldovich ($n=1$) neutrino¹⁶ and cold DM spectra. b , r.m.s. mass fluctuation within a randomly placed sphere containing mass M for cold DM, $n=1$, and ($\Omega=1$, $h=0.5$), ($\Omega=0.2$, $h=1$).

invisible—for example, in the form of diffuse ionized intergalactic gas at $T \sim 10^4$ K.) This is illustrated in Table 1 and Fig. 1, where we have plotted observed data for M/L and M/M_{lum} . The fact that the total-to-luminous mass of rich clusters is similar to that of galaxies including their massive haloes, even though the clusters' mass-to-light ratio is larger, is due mainly to the different stellar population in the E and S0 galaxies in rich clusters plus the large contribution of X-ray emitting gas to M_{lum} . In very rich clusters such as Coma, there is perhaps ~ 2 – 5 times as much mass in hot gas as there is in stars.

Finally, preliminary velocity dispersion data for Draco, Carina, and Ursa Minor^{48–50} as well as theoretical arguments⁵¹ suggest that a significant amount of DM may reside in dwarf spheroidal galaxies. Because of the low velocity dispersion of dwarf galaxies, phase space constraints give a lower limit of $m > 500$ eV for the mass of particles comprising this DM³². The present velocity dispersion estimates are uncertain owing to possible stellar oscillations, mass outflow, and binary motions, but these effects can be discovered and eliminated with careful monitoring. Taken at face value, however, the existing mass limit of 500 eV would rule out neutrinos completely.

Although warm DM provides a natural (free streaming) scale for ordinary galaxies, it cannot account for massive haloes in dwarf spheroidals even though the warm DM mass is actually, if barely, consistent with the phase space constraint. This is because free streaming damps out fluctuations with mass $< 10^{11} M_{\odot}$, so dwarf galaxies with mass $\sim 10^7 M_{\odot}$ can form in this picture only by the fragmentation of much larger scale galactic masses. Because dwarf galaxies with escape velocity $\sim 10 \text{ km s}^{-1}$ would capture only a small fraction of the warm DM particles, whose velocity dispersion would be $\sim 100 \text{ km s}^{-1}$, typical of ordinary spirals, one would expect M/M_{lum} to be much smaller in dwarf galaxies than in spirals. However, the dynamical mass data on dwarf spheroidals shown in Fig. 1 suggest a similar value for the ratio of total-to-luminous mass. In view of the great importance of dwarf spheroidal haloes to constraints on warm as well as hot DM, it is clearly urgent to continue the monitoring programme on velocity dispersions mentioned above.

Recently proposed models having two or more kinds of possibly unstable DM^{52–58} may be easier to reconcile with observations than models with one stable DM species, in part because of their additional adjustable parameters. Such models are, however, beyond the scope of this paper.

Galaxy formation with cold dark matter

The key features of galaxy formation in the cold DM picture are these: the DM fluctuation spectrum at recombination is determined by a small number of physical parameters; after recombination (at $z_{\text{rec}} \approx 1,300$) the amplitude of the baryonic fluctuations rapidly grows to match that of the DM fluctuations; smaller-mass fluctuations grow to nonlinearity and virialize, and then are hierarchically clustered within successively larger bound systems; and finally the ordinary matter in bound systems of total mass $\sim 10^{8-12} M_{\odot}$ cools rapidly enough within the DM haloes to form galaxies, while larger mass fluctuations form clusters.

To calculate the growth of initially small perturbations and their ultimate gravitational collapse into condensed systems, one must first determine an initial spectrum for density perturbations in the very early Universe. We assume here that the initial fluctuations are adiabatic, which is consistent with fashionable particle theories in which a small excess of baryons over antibaryons is generated by the decay of supermassive grand unified theory particles^{40,41}. Then the mass (or energy) density at any point can be written as $\rho(r, t) = \rho_0(1 + \delta)$, where $\rho_0(t)$ is the average density in the Universe and δ represents the fractional density perturbation in the synchronous gauge. If the fluctuation spectrum is characterized by a power law distribution, then the r.m.s. fluctuation on mass scale M can be written $\delta \propto M^{-(1/2) - (n/6)}$, which corresponds to a Fourier power spectrum $|\delta_k|^2 \propto k^n$. Alternatively, we can characterize the perturbations

Table 1 M/L_B and M/M_{lum} on various mass scales

Unit	M^*	$M/L_B^\dagger$	$M_{\text{gas}}/M_{\text{lum}}$	M/M_{lum}
Large clusters	$10^{15} M_{\odot}$	$316 \pm 40^\ddagger$	$0.84^{+0.0}_{-0.1}§$	$8.4^{+7.0}_{-1.0}$
Small E-dominated groups	5×10^{13}	$83^{+80}_{-10} $	$0.61^{+0.1}_{-0.1} $	$5.4^{+10.0}_{-2.0} $
Small spiral-dominated groups	2×10^{13}	$40^{+50}_{-10}¶$	$0(?)$	$14.2^{+36}_{-6} \#$
Whole Milky Way	$10^{12} **$	$50^\ddagger$	$0(?)$	$14^\ddagger$
Dwarf spheroidals§§				
Stellar masses	10^{5-7}	2.5	0	1
Dynamical masses	10^{6-8}	30	0	12

* Total mass including dark matter. $H = 50 \text{ km s}^{-1} \text{ Mpc}$.

† Mass-to-light ratio on B(0) magnitude system as described in ref. 1.

‡ From ref. 120.

§ $M_{\text{gas}}/M_{\text{lum}}$ is calculated as the product of two factors: $(M_{\text{gas}}/M) \times (M_{\text{lum}}/M)^{-1}$. $M_{\text{gas}}/M = 0.10^{+0.02}_{-0.05}$ from ref. 121 for 14 X-ray clusters. Formal clusters have been increased to allow for the unknown gas distribution outside the core and possible settling of gas relative to galaxies or vice versa. M_{lum}/M is $(M_*/M_{\text{gas}})/M$, where M_*/M is the fractional mass in stars. It may be calculated as $(M/L_B)_*/(M/L_B)^{-1}$, where $(M/L_B)_*$ is the stellar mass-to-light ratio. For E and S0 galaxies in large clusters, $(M/L_B)_*$ appears to be about 6 (ref. 1).

¶ L_V , M_{gas} , and M from ref. 122 based on five groups. L_V corrected to B(0) system based on data provided by Kriss (personal communication). Comparison with Beers *et al.*¹²³ suggests that M may be underestimated by a factor of two. Quoted errors take this into account.

¶ From ref. 1.

From ref. 72; includes baryonic mass in stars and neutral gas.

** Mass assumed to be 1/3 total mass of Local Group¹.

†† B(0) luminosity from ref. 1.

‡‡ M_{lum} from ref. 124.

§§ Basic data from ref. 51. Upper line assumes that only stars are present. Lower line includes excess dark matter, based on mean dynamical masses of Faber and Lin⁵¹ and Aaronson *et al.*^{48–50}.

on mass scale M when that mass scale crosses (first comes within) the horizon as $\delta_H = \epsilon (M/M_0)^{-\gamma}$, where M_0 is the present horizon mass and $n = 6\gamma + 1$. To avoid having too much power on either large or small scales requires $\gamma \approx 0$, which corresponds to $n \approx 1$. Limits on the large-scale variation of the microwave background further imply $\epsilon \leq 10^{-4}$. The case $n = 1$ is commonly referred to as the constant curvature or (Harrison–Peebles–) Zeldovich^{59–61} spectrum, and is predicted in inflationary models. It has recently been shown that not only inflation but also $\epsilon \approx 10^{-4}$ can be arranged in suitably fine-tuned grand unified theories^{62–64}. The Zeldovich spectrum also arises automatically if the fluctuations are due to cosmic strings^{65,66}. We show below that $n = 1$ is, furthermore, in best agreement with galaxy and cluster data.

The first study of the growth of cold DM fluctuations was the numerical calculation of Peebles², who included cold DM, photons, and ordinary matter. Two of us (G.R.B. and J.R.P.)^{3–5,7} extended these numerical calculations to include three massless neutrino species. Fluctuations having mass $M < M_{\text{eq}} = 2 \times 10^{15} (\Omega h^2)^{-2} M_{\odot}$ cross the horizon when the Universe is still radiation dominated, that is, when $z > z_{\text{eq}} = 2.5 \times 10^4 \Omega h^2$. After such fluctuations cross the horizon, their neutrino components dissipate by free streaming, and the remaining photon and charged particle perturbations oscillate as an acoustic wave (whose amplitude is ultimately damped by photon diffusion for⁶⁷ $M < M_{\text{Silk}} \approx 3 \times 10^{13} \Omega_b^{-1/2} \Omega^{-3/4} h^{-5/2} M_{\odot}$). As a result, the main driving terms for the growth of cold dark matter fluctuations δ_{DM} decrease, and consequently δ_{DM} begins to grow only very slowly until the Universe becomes (dark) matter dominated at z_{eq} , after which $\delta_{\text{DM}} \propto a = 1/(1+z)$ until $z \approx \Omega^{-1}$. This stagnation of the growth of DM fluctuations between the epochs of horizon crossing and matter domination is called 'stagspansion'^{3,7}.

Because fluctuations with $M < M_{\text{eq}}$ grow very little during the stagspansion era and as fluctuations on all scales grow at essentially the same rate after the Universe becomes matter dominated, an initial Zeldovich spectrum, $\delta_{\text{DM}} \propto M^{-2/3}$, evolves to a much flatter spectrum for $M < M_{\text{eq}}$ by the time of recombination. In Fig. 2 we plot the resulting cold DM fluctuation spectrum at present (ignoring the nonlinear evolution which is important for $\delta \geq 1$, as discussed below; thus the plotted spectrum δ_k is larger than the spectrum at recombination by a constant factor). The quantity plotted in Fig. 2a, $k^{3/2} |\delta_k|$, roughly

represents the spatial density fluctuation δ as a function of total mass $M = 4\pi^4 \rho_0 / 3k^3$. Also sketched are the fluctuation spectra in the hot DM scenario and in an isothermal scenario with white noise fluctuations. Note that the hot DM model requires more power on large scales today to form superclusters by $z \approx 2$. Figure 2b shows cold DM fluctuation spectra plotted in another form, $\delta M/M$, which represents the r.m.s. mass fluctuation within a randomly-placed sphere of radius R containing mass M . Following Peebles², we have normalized the curves so that, at the present epoch, $\delta M/M = 1$ at $R = 8h^{-1}$ Mpc.

Even though baryonic fluctuations do not grow (and are damped for $M < M_{\text{Silk}}$) before recombination, after recombination the baryons 'fall into' the DM perturbations so that quickly $\delta_b = \delta_{\text{DM}}$ (refs 68, 69). This will occur if $\Omega_{\text{DM}} \delta_{\text{DM}} \gg \Omega_b \delta_b$, as we expect, and if the baryonic fluctuation mass exceeds the baryonic Jeans mass⁶ $M_{J,b} \sim 10^6 \Omega_b \Omega^{-3/2} h^{-1} (T_b/T)^{3/2} M_\odot$, where T_b is the temperature of the baryonic gas and T is the photon temperature. On scales smaller than this, the pressure of the baryonic gas prevents it from developing the same density contrast δ as the cold DM. The value of T_b/T is kept close to unity for $z > 100$ by the coupling of the residual free electrons with the radiation, but for smaller z it falls off approximately as $(1+z)$. This means that at $z \approx 10$, $M_{J,b}$ may be as small as $\sim 10^3 M_\odot$.

At any mass scale M , when the fluctuation $\delta M/M$ approaches unity, nonlinear gravitational effects become important. The fluctuation then separates from the Hubble expansion, reaches a maximum radius, and begins to contract. Spherically symmetric fluctuations, for example, contract to about half their maximum radii. During this contraction, violent relaxation due to the rapidly varying gravitational field converts enough potential energy (PE) into kinetic energy (KE) for the virial theorem, $\langle PE \rangle = -2\langle KE \rangle$, to be satisfied. After virialization, the mean density within a fluctuation is roughly eight times the density corresponding to the maximum radius of expansion⁸.

As the cold-DM fluctuation spectrum $\delta M/M$ is a decreasing function of M , smaller mass fluctuations will, on the average, become nonlinear and begin to collapse at earlier times than larger mass fluctuations. Smaller mass fluctuations are themselves typically clustered within larger mass perturbations, which go nonlinear later. This hierarchical clustering of smaller systems into larger and yet larger gravitationally bound systems begins at the baryon Jeans mass, $M_{J,b}$, and continues until the present. The baryonic substructures within larger mass clusters will then be disrupted by virialization of the clusters unless significant mass segregation between baryons and DM has occurred before cluster virialization. Hence, to maintain their existence as a separate substructure, the baryons must cool and gravitationally condense within their massive DM haloes before virialization occurs on larger scales⁷⁰.

Figure 3 shows the baryonic number density n_b plotted against temperature T (or halo velocity dispersion) for virialized spherically symmetric protocondensations resulting from an initial Zeldovich spectrum of cold DM fluctuations. The simplifying assumption of sphericity is not expected to be an unreasonable approximation for hierarchical clustering. Two cases are considered: $\Omega = 1$ and $h = 0.5$ (solid lines in Fig. 3), and $\Omega = 0.2$ and $h = 1$ (dashed lines). The curves assume that the protocondensations have virialized, but that the baryons have not yet cooled and condensed. The curves labelled 1σ assume that for each mass scale, $\delta M/M$ has the r.m.s. value; those labelled 2σ correspond to fluctuations $\delta M/M$ twice as great; and so on. For given total mass M , the distribution of values of $\delta M/M$ spreads out along lines of constant M (light diagonal lines in Fig. 3). We have also shown in Fig. 3 the present positions of clusters and groups of galaxies and of individual galaxies, including dwarf spheroidals. Note that different types of galaxies, the Hubble sequence, are spread out in this diagram.

Baryonic cooling moves structures downward in Fig. 3, below the model curves. Baryons can radiatively cool through collisional excitation of atoms and molecules, and by Compton cooling off the cosmic background radiation; Fig. 3 correspondingly includes the medium solid curves labelled 'No metals' and

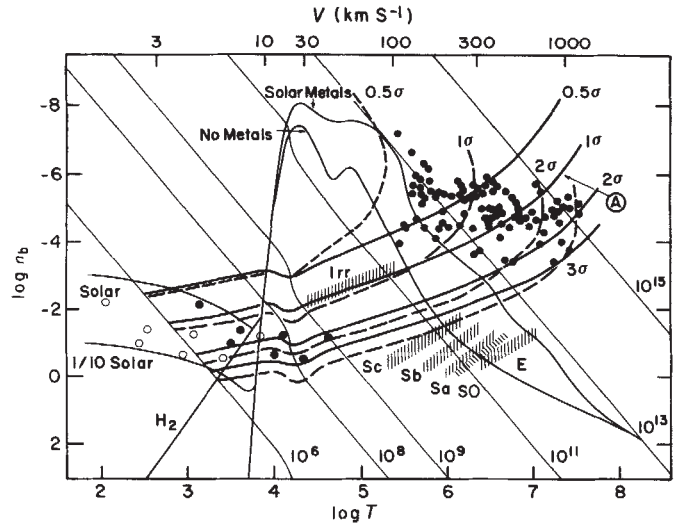


Fig. 3 Baryon density n_b versus three-dimensional, r.m.s. velocity dispersion V and virial temperature T for structures of various size in the Universe. The quantity T is $\mu V^2/3k$, where μ is mean molecular weight (≈ 0.6 for ionized, primordial H + He) and k is Boltzmann's constant. Dots at upper right represent nearly all groups and clusters within 5,000 km s⁻¹ in the CfA catalogue⁸². Baryon density is obtained from total density assuming $\rho_b/\rho = 0.1$ (see Table 1), where ρ is defined as $M/(4\pi R_{\text{vir}}^3/3)$, M is total mass, and R_{vir} is GM/V^2 . Catalogue groups are chosen to exceed a minimum threshold luminosity density. Their minimum baryon density should thus fall on a horizontal line, slightly modified by the higher M/L of rich clusters, as is observed. Point A represents Abell clusters of richness 2 and 3 studied by Dressler¹²⁰. Here $R_{\text{vir}} = 3.0 R_{\text{eff}}$, valid for a de Vaucouleurs cluster profile¹²⁵. The quantity T for clusters corresponds to central velocity dispersion, with the arrow indicating the effect of the falloff in velocity dispersion at large radii observed in the Coma cluster^{126,127}. For galaxies, M_{lum} is set equal to $(M_{\text{lum}}/L_B) \times L_B$, where L_B is blue luminosity. Assumed values of (M_{lum}/L_B) : E = 8, S0 = 6, Sa = 4, Sb = 2.5, Sc = 1.5, Dw = 1.00, D = 0.5. Spheroidal = 2.5 (all but the last are based on $h = 0.5$). Galaxy radii are approximate virial radii, assuming that the baryonic components of galaxies are self-gravitating (to obtain a radius consistent with the definition used for groups and clusters above). To achieve this, isophotal radii from various sources have been appropriately scaled, as follows: E-Sb, $R_{\text{vir}} = R_{25}$ (refs 128-131); Sc, $R_{\text{vir}} = 1.24 R_{25}$ (ref. 132); Dw-Irr, $R_{\text{vir}} = 1.24 R_H$ (ref. 133); Dw-Sph, $R_{\text{vir}} = 0.70 R_{\text{id}}$ (ref. 51). Velocities: E-S0, $V = \sqrt{3} \times \sigma$ (nucleus) (refs 129, 134); Sa-Sc, $V = v_{\text{rot}}^{\text{max}}$ (refs 130, 131, 132); Dw-Irr, $V = 1/2$ FWHM (Δv) (ref. 133), corrected for inclination. Dw Spheroidals are plotted twice; open circles, $M = 2.5 L_B$ (stars only) and $V = (GM/R_{\text{vir}})^{1/2}$; large dots, $M = 30 L_B$ (stars plus dark matter⁴⁸⁻⁵¹), with V as before. Light diagonal lines represent the masses of self-gravitating bodies with the indicated values of n_b and T and assuming $\rho_b = 0.1\rho$. The discontinuities near 10^4 K are due to the effects of H and He ionization. Cooling curves^{78,135,136} (medium lines) separate regions where cooling is efficient ($\tau_{\text{cool}} < \tau_{\text{dyn}}$, lower region) from regions where it is inefficient ($\tau_{\text{cool}} > \tau_{\text{dyn}}$, upper region). All curves assume a residual electron fraction after recombination of 10^{-4} , consistent with $\Omega_b/\Omega = 0.1$. Model curves represent the equilibria of structures that collapse dissipationlessly from the cold dark matter initial fluctuation spectra of Fig. 2 with $n = 1$. DM haloes and groups and clusters of galaxies should lie on these curves, whereas the baryon components of galaxies should lie below due to baryonic dissipation. The curves labelled σ refer to fluctuations with $\delta M/M$ equal to σ times the r.m.s. value shown in Fig. 2. Heavy curves: $\Omega = 1$, $h = 0.5$; dashed curves: $\Omega = 0.2$, $h = 1$. Model curves appear to pass through groups and clusters in about the right place, and the horizontal spread in T is roughly as expected for a gaussian distribution in $\delta M/M$. Baryon components of galaxies lie below the loci for dissipationless collapse and generally within the region where strong baryonic cooling is expected. Dwarf spheroidal galaxies lie near the beginning of the clustering hierarchy and may be typical of the earliest structures to collapse (see ref. 73). Hubble types are spread out along different loci, perhaps due in part to different baryon collapse factors (Es larger, Irrs smaller) and in part to intrinsic differences in initial $\delta M/M$ (see also Fig. 4).

'Solar metals', 'H₂', and 'Compton'. Below these curves the baryonic cooling time is shorter than the dynamical time and above them the reverse is true⁷¹. Figure 3 immediately shows that, while the Hubble sequence of galaxies shows strong evidence for baryonic cooling and dissipation (core condensation in heavy haloes), dwarf spheroidals are only marginally able to cool, and groups and clusters of galaxies have too long a cooling time to have dissipated much energy on their scale^{72,73}.

On average, for the cold DM fluctuation spectrum, what range of total masses yields baryon condensation within a massive halo? Let us first consider large galaxies. Naively, Fig. 3 suggests

that there is an upper bound for galactic masses of $M \leq 10^{12} M_{\odot}$, where the baryonic cooling time for gas of primordial composition begins to exceed the dynamical time. More massive galaxies could form in dense regions, where perhaps an early generation of Population III stars enriched the remaining baryonic gas with metals. Actually, the situation is likely to be more complicated, with many large galaxies formed by the merger of smaller ones. The resulting galaxy mass distribution arises from competition between hierarchical clustering and decreasing galaxy collision cross-sections due to dissipation, and it is thus difficult to calculate reliably. Regarding the smallest galaxies, collisional excitation of atomic hydrogen provides a lower limit of $M \geq 10^8 M_{\odot}$, corresponding to virialized baryonic temperature $T_b \geq 10^4$ K. This range of protogalactic total masses, $10^8 M_{\odot} \leq M \leq 10^{12} M_{\odot}$, encompasses virtually all the mass that is observed to comprise galaxies. For protogalaxies in this mass range, the velocity dispersion of the baryons will initially remain nearly constant ($T \approx \text{constant}$) as they condense within the gravitational potential of the virialized (and presumably roughly isothermal) DM halo. When the baryon density increases enough that their gravitational potential dominates that of the halo, the baryons' velocity dispersion will rise as they continue to dissipate energy and condense. Baryonic contraction is finally halted by rotation and perhaps, in some protogalaxies, by star formation (see below).

The collapse of fluctuations with mass $M > 10^{13} M_{\odot}$ leads to clusters of galaxies in this picture. In clusters, only the outer parts of member galactic haloes are stripped off by collisions—the inner baryonic cores are able to contract to form the observed stellar systems. More of the baryons in the richest clusters are observed to be in the form of hot gas than in galaxies, as we have already mentioned in connection with Fig. 1. Perhaps this is because rich clusters tend to contain high-density cores, which collapse early, simultaneously with many of the galaxies they contain.

Can the cold DM picture account for the wide range of morphologies displayed by clusters of galaxies in X-ray⁷⁴ and optical-band⁷⁵ observations, ranging from regular, apparently relaxed configurations to complex, multicomponent structures? Preliminary results are encouraging. In particular, simulations show that large central condensations form quickly and can grow by subsequent mergers to form cD galaxies if most of the DM is in haloes around the baryonic substructures, as expected for cold DM, but not if the DM is distributed diffusely^{76,77}.

What happens to small clouds whose baryonic mass M_b lies in the range $M_{J,b} < M_b < 10^{7-8} M_{\odot}$ and for which $T_b < 10^4$ K after virialization? For a primordial element abundance, the molecular cooling time (primarily due to H_2) is less than the dynamical time for fluctuations that satisfy

$$M_b > M_{C,b} \approx 4.1 \times 10^6 (\Omega h^2)^{-0.917} \left(\frac{Y_e}{10^{-4}} \right)^{-0.625} \left(\frac{\Omega_b}{0.1\Omega} \right)^{-1.04} \\ \times \left(\frac{1+z_t}{10} \right)^{-2.75} M_{\odot}$$

where z_t is the redshift of maximum expansion of the fluctuation. (This result uses the cooling rates given by Yoneyama⁷⁸ and assumes the simple 'top hat' model of collapsing fluctuations. It corrects an equation given in ref. 17.) The quantity Y_e , the fraction of free electrons, is essentially the fraction that escapes recombination as the universe cools below $\sim 1,000$ K, and is roughly $10^{-4}(\Omega_b/0.1)^{-1}\Omega^{1/2}h^{-1}$. Systems with $M_b < M_{C,b}$ will persist as pressure supported clouds until they are disrupted by the virialization of larger scale clusters. Clouds having $M_b > M_{C,b}$ can dissipate energy and collapse, although the influence of rotation and the efficiency of fragmentation are poorly understood. The end product may be an irregular or dwarf spheroidal galaxy. Other possibilities such as a protoglobular cluster or one or more very massive objects (VMOs) would require greater contraction and are likely to be inhibited by angular momentum.

Moreover, it is unclear what fraction of the original baryonic mass can be retained by such small clouds rather than expelled.

There is—despite the uncertainties—a strong probability that energy output from fragmented subsystems can influence the gas that remains uncondensed, thereby exerting a feedback on further condensation. In particular, UV emission from massive or supermassive stars can photoionize the remaining diffuse baryons, raising T_b up to 10^4 K. If such stars radiate a fraction q_{UV} of their Eddington luminosity in Lyman continuum photons, then the entire baryonic medium can be ionized even by a mass fraction of stars as low as $\sim 3 \times 10^{-5}/q_{UV}$. If the baryonic gas is re-ionized in systems with total mass $M \leq 10^8 M_{\odot}$, the baryons become so hot that they flow out of the cloud, whose gravitational field is not strong enough to bind them.

This may explain the origin of dwarf spheroidal galaxies, with $10^6 M_{\odot} < M < 10^8 M_{\odot}$, which show little evidence of dissipation in Fig. 3. As the baryons in these systems begin to condense through molecular cooling, if fragmentation and star formation are very efficient the baryons quickly turn into stars and condensation ceases. On the other hand, if a fraction of the baryons fragment into stars having strong UV luminosity, the remaining baryons are heated until they leave the system, and dissipation will again cease. The apparent absence of dissipation in dwarf spheroidals could also be due to stripping of their baryons in encounters with more massive systems³², which would move the dwarfs upward in Fig. 3.

In the cold DM scenario, globular clusters are probably not primordial objects. For one thing, there is no evidence that they have massive haloes (although, as Peebles⁷⁹ points out, there is not much direct contrary evidence either). Furthermore, if truly primordial, they should be distributed in the Universe like DM whereas, at least within galaxies, they seem to have dissipated and condensed like the other baryonic matter. However, the existence of 'standard objects' inside galaxies with mass $\sim 10^6 M_{\odot}$ demands some explanation. In the model discussed here, there is a natural mass scale of order $M_{g,b} \sim M_{Gal,b}(T_{virial}/10^4 \text{ K})^{-2}$, where $M_{Gal,b}$ is the baryonic mass within a galaxy; $M_{g,b}$ is the Jeans mass of a cloud at 10^4 K in pressure balance with protogalactic gas at the virial temperature⁸⁰. During the dissipation phase of galaxy formation, the gas might be likely to have a two-phase structure with a hot phase at T_{virial} and a cool phase at $\sim 10^4$ K, in which case subcondensations of mass $M_{g,b}$ would be expected with density contrast $\approx T_{virial}/10^4$ K. We identify these with protoglobular clusters (noting that the considerations of Fall and Rees⁸¹ could further limit the mass range).

Galaxies and clusters

While the n_b - T plot (Fig. 3) is useful for comparing data and predictions with the cooling curves, it is also useful to consider total mass M versus T , as in Fig. 4. This avoids having to take into account the differing amounts of baryonic dissipation suffered by various galaxies. The heavy solid and dashed curves again correspond to the $n=1$ cold DM spectrum, for $(\Omega=1, h=0.5)$ and $(\Omega=0.2, h=1)$ respectively. It is striking that galaxies in the M - T diagram lie along lines of roughly the same slope as these curves. This occurs because the effective slope of the $n=1$ cold DM fluctuation spectrum in the galaxy mass range is $n_{eff} \approx -2$, which corresponds to the empirical Tully-Fisher and Faber-Jackson laws: $M \propto v^4$. The light dashed lines in Fig. 4 are the postvirialization curves for primordial fluctuation spectra with $n=0$ (white noise) and $n=2$. The $n=1$ (Zeldovich) spectrum is evidently the one that is most consistent with the data.

The points in Fig. 4 represent essentially all of the clusters identified by Geller and Huchra⁸² in the CfA catalogue within $5,000 \text{ km s}^{-1}$. The cluster data lie about where they should on the diagram, and even the statistics of the distribution seem roughly to correspond to the expectations represented by the $0.5, 1, 2$, and 3σ curves. Note that the galaxies and clusters are neatly separated by the Compton cooling line of Fig. 3, suggesting that the era of galaxy formation ceased in the Universe when Compton cooling off microwave background photons became inefficient (J. E. Gunn, personal communication).

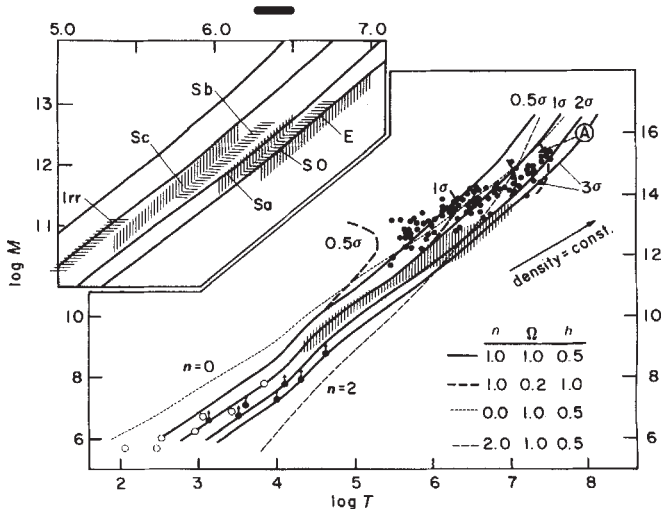


Fig. 4 Total mass M versus virial temperature T . Data sources and symbols are as in Fig. 3. M for groups and clusters is total dynamical mass. For galaxies, M is assumed to be $10 M_{\text{lum}}$ (see Fig. 3 legend). If Dw spheroidals actually have $M/L_B = 30$, they may have suffered baryon stripping³², in which case M is a lower limit (arrows). Details of the region occupied by massive galaxies are shown in the inset, upper left. In addition to the $n=1$ models from Fig. 3, two 1σ curves for $n=0$ and $n=2$ are also shown (light dashes). Either set of curves for an $n=1$ (Zeldovich) spectrum provides a good fit to the observations over 9 orders of magnitude in mass. Curves with $n=0$ and $n=2$ do not fit as well. The apparent gap between galaxies and groups and clusters in Fig. 3 (which stems from baryonic dissipation) vanishes in this figure, and the clustering hierarchy is smooth and unbroken from the smallest structures to the largest ones. The Fisher-Tully and Faber-Jackson laws for galaxies ($M \propto V^4$ or T^2) arise naturally as a consequence of the slope of the cold DM fluctuation spectrum in the mass region of galaxies. Groups and clusters are distributed around the $n=1$ loci about as expected. The apparent upward trend among the groups is not physically meaningful but arises from their selection as density enhancements above a minimum threshold (see caption, Fig. 3, and constant-density arrow, this figure). The exact locations of galaxies are uncertain. In particular, the temperatures of Es and S0s may be overestimated owing to the use of nuclear rather than global velocity dispersions. The masses of Dw Irrs may also be too low owing to neglect of mass in H_2 . Taken at face value, however, the data suggest that early-type galaxies (Es and S0s) arise from high- $\delta M/M$ fluctuations, whereas late-type galaxies (Scs and Irrs) arise from low- $\delta M/M$ fluctuations. Groups and clusters appear to fill a wider band than galaxies. If real, this difference may indicate that low- $\delta M/M$ fluctuations on the mass scale of galaxies once existed but did not give rise to visible galaxies. This suggests further that galaxy formation, at least in some regions of the Universe, may not have been fully complete and that galaxies are therefore not a reliable tracer of total mass. The sharp demarcation line between galaxies and clusters corresponds to the Compton cooling limit in Fig. 3 and suggests that galaxy formation ceased when Compton cooling off the microwave background became inefficient.

Note that spiral galaxies lie roughly along the 1σ curve while elliptical galaxies lie along the 2σ curve. Although this displacement is not large compared with the uncertainties, it is consistent with the fact that more than half of all galaxies are spirals, while only $\sim 15\%$ are ellipticals. We have suggested elsewhere⁸³ that, in hierarchical clustering scenarios, the higher σ fluctuations will develop rather smaller angular momenta, as measured by the dimensionless parameter $\lambda (=JE^{1/2}G^{-1}M^{-5/2})$. This difference seems to exist with either white noise or a flatter spectrum, but to be larger in the latter case. If high σ fluctuations have little angular momentum, their baryons can collapse by a large factor in radius, forming high-density ellipticals and spheroidal bulges, as shown in Fig. 3. Because, with a flat spectrum, higher σ fluctuations occur preferentially in denser regions destined to become rich clusters (the statistics of such correlations can be treated⁸⁴ by the methods of Peebles⁶), one expects to find more ellipticals there—as is observed. Indeed, the rich clusters lie along the same 2 and 3σ curves in Fig. 4 as do the elliptical galaxies.

Presumably the collapse of the low- λ protoelliptical galaxies is halted by star formation well before a flattened disk can form, yielding a stellar system of spheroidal shape. The mechanism governing the onset of star formation in these systems is unfortunately not yet understood, but may involve a threshold effect which sets in when the baryon density exceeds the DM halo

density by a sufficient factor^{70,72}. Disks (spirals and irregulars) form from more typical higher- λ protogalaxies, which, for a given mass, are larger and more diffuse than their protoelliptical counterparts. The collapse of disks thus occurs through a relatively slow infall of baryons from $\sim 10^2$ kpc, and is halted by angular momentum. Infall from such distances is consistent both with the extent of dark haloes inferred from observations and with the high angular momenta of present-day disks ($\lambda \sim 0.4$)^{85,86}. The location of galaxies in Fig. 3 is consistent with these ideas if the baryons in all galaxies collapsed in radius by roughly the same factor, about an order of magnitude, but less for late-type irregulars and more for early-type Es and spheroidal bulges.

It has been theorized that the Hubble sequence originates in the distribution of either the initial angular momenta^{87,88} or else the initial densities⁸⁹ of protogalaxies. However, if overdensity and angular momentum are linked, with the high- σ fluctuations having lower λ , then these two apparently competitive theories become opposite sides of the same coin.

Consider finally the difference in Fig. 4 between the solid and dashed lines. The dashed lines, representing a lower-density Universe ($\Omega = 0.2$), curve backward at the largest masses and lie far away from the circle representing the cores of the richest clusters, Abell classes 2 and 3. Because these regions of very high galaxy density contain at least several per cent of the mass in the Universe, the circle should lie between the 2 and 3σ lines (assuming gaussian statistics). It does so for the solid ($\Omega = 1$) lines, but not for the dashed lines. At face value, this is evidence favouring an Einstein-de Sitter Universe for cold DM. However, there are at least two reasons why this argument should not be taken too seriously. First, the velocity dispersions represented by the Abell cluster circle in Fig. 4 correspond to the cluster cores. The model curves on the other hand refer to the entire virialized cluster, over which the velocity dispersion is considerably lower (as indicated schematically by the arrow attached to the circle in Fig. 4). Second, the assumption of spherical symmetry used in obtaining both sets of curves is only an approximation. The initial collapse is probably often quite anisotropic—more like a Zeldovich pancake than a sphere. It is, therefore, preferable to compare these data with N -body simulations than with the simple model represented by the curves in Fig. 4. Until this becomes possible we do not believe that the data in Fig. 4 allow a clear-cut discrimination between the $\Omega = 0.2$ and $\Omega = 1$ cases, especially if the Hubble parameter h is allowed to vary simultaneously within the observationally allowed range, as we have assumed.

Cosmological density

The most straightforward interpretation of the approximate constancy of $M/M_{\text{lum}} \approx 10$ from galaxy through rich cluster scales (Fig. 1) is that dark matter clusters with galaxies. This is precisely what cold DM is expected to do. One then expects the density parameter $\Omega = (M/M_{\text{lum}})\Omega_{\text{galaxies}} \approx 10 \times 0.02 = 0.2$, in agreement with $\Omega = 0.2 \times 1.5^{\pm 1}$ from the cosmic energy equation and the stability of clustering⁹⁰.

The value $\Omega \approx 0.2$ is consistent with all reliable measurements and furthermore gives an age for the Universe, t_0 , consistent with globular cluster age estimates⁹¹ (≥ 15 Gyr) if $h \leq 0.6$. The observed abundance of deuterium plus helium-3 implies a lower limit³⁷ $\Omega_b h^2 \geq 0.01$, which is also consistent with $\Omega = 0.2$ and $\Omega_b \approx 0.02$ if $h \geq 0.7$.

Another argument favouring $\Omega = 0.2$ for cold DM is based on preliminary N -body results⁹², which indicate that superclusters and voids form on the observed scale for $\Omega = 0.2$, but on too small a scale for $\Omega = 1$ unless the Hubble constant is unrealistically small. However, this conclusion follows from comparing the N -body mass autocorrelation function $\xi_m(r)$ with the observed galaxy autocorrelation function $\xi_g(r)$. This is justified only if the galaxies are a good tracer of mass—if $M/L \approx$ constant. This is certainly not true for rich clusters, as illustrated in the top part of Fig. 1 and in Table 1, where M/L for very rich clusters is roughly six times larger than it is for small groups

and for the Milky Way (including its DM halo). This means that some of the actual small-distance-mass autocorrelation is not included in the galaxy autocorrelation function; that $\xi_m(r)$ is steeper for $r \leq 1$ Mpc than $\xi_g(r)$. This effect is in the right direction to bring the $\Omega = 1$ N -body simulations into consistency with observations. Note that galaxy velocity dispersion averages also underestimate the actual mass-weighted velocity dispersion, because the high-velocity, dense regions of rich clusters are under-represented in the counts. This effect is also compatible with higher Ω .

The absence of measurable fluctuations in the cosmic microwave background also restricts Ω , because in a low- Ω Universe there is less growth of fluctuations both before and after recombination. The latest limits^{38,39} on $\Delta T/T$ require $\Omega \geq 0.2h^{-4/3}$ (refs 93, 94), unless the Universe was reionized after recombination.

Both prejudice and the inflationary hypothesis favour the Einstein-de Sitter value $\Omega = 1$. (Actually, inflation implies more generally that $\Omega = 1 - \Lambda/3H^2$, where H is Hubble's constant and Λ is the cosmological constant⁹⁵, but we assume here that $\Lambda = 0$.) For $\Omega = 1$, $t_0 = 6.5h^{-1}$ Gyr, so consistency with global cluster age estimates requires the Hubble constant to be perhaps unrealistically small. As we have discussed, the Zeldovich ($n = 1$) spectrum of primordial adiabatic fluctuations, which also follows from inflation, is compatible with models of galaxy formation in the cold DM scenario. Of course, the Zeldovich spectrum does not necessarily entail inflation. Moreover, inflation does not necessarily imply $\Omega = 1$: if there is a great deal of inflation, then, of course, Ω is very close to unity (assuming vanishing cosmological constant); but if we speculate that greater amounts of inflation are increasingly unlikely, then our horizon might happen to lie in a patch with $\Omega = 0.2$.

Could $\Omega = 1$ in the cold DM model? Based on Fig. 1, this could happen only if M/M_{lum} increases substantially on scales larger than rich clusters. Perhaps galaxy formation is suppressed in the voids, and the resulting luminosity contrast is so strong, for reasons we do not yet understand, that $M/L > 1,500$ there (see below). Alternatively, if the virial mass of clusters of galaxies is significantly underestimated, as suggested recently by Kaiser, and if galactic haloes extend much further than ~ 70 kpc, then the ratio M/M_{lum} may be substantially underestimated in Fig. 1, leading to a larger value for Ω . If galaxy formation is inefficient so that only large-overdensity perturbations can form galaxies—a possibility suggested by Fig. 4—the $\Omega = 1$, cold DM N -body simulations⁹² can be brought into agreement with observations^{96,97}. We conclude that a straightforward interpretation of the evidence summarized above favours $\Omega \approx 0.2$ in the cold DM picture, but that $\Omega = 1$ is not implausible.

Note that the cold DM model may require some means of suppressing galaxy formation in voids for $\Omega \approx 0.2$ as well as for $\Omega = 1$. This is required in the latter case to hide most of the mass. In a low- Ω Universe, on the other hand, large, very underdense regions cannot form by gravitation alone^{98,99}; and galaxy formation must be suppressed somehow in regions of moderately low density if the density of galaxies in voids is less than one-quarter of the average density, the quoted upper limit for the Boötes void^{100,101}. Suppression of galaxy formation may occur partly because of the substantial difference at $z < \Omega^{-1}$ between $\langle \rho \rangle$ and ρ_c in a low-density Universe.

Superclusters and voids

Recent, accurate redshift measurements of several thousand galaxies have revealed the presence of voids with linear dimensions ranging up to $\sim 50h^{-1}$ Mpc that are almost completely empty of bright galaxies. Most galaxies are concentrated in irregularly shaped, flattened, or elongated superclusters (see ref. 102 for fuller discussion and references). Two related observations are particularly interesting for their possible bearing on the origin of this large scale structure: a correlation of galaxy type with galaxy number density^{103,104} and, possibly, a correlation between the orientation of cluster major axes and the direction to neighbouring clusters within $\sim 15h^{-1}$ Mpc (ref. 105).

In the hot DM picture, the fluctuation spectrum is rather sharply peaked at masses $\sim 4 \times 10^{15}$ (30 eV/ m)². The suppression of smaller scale structure results in an essentially pressure-free collapse on the supercluster scale, with the formation of sharp caustics—Zeldovich pancakes^{106,107}. If galaxies form only in the densest regions, where the shocked baryons have cooled sufficiently^{108,109}, most of the Universe would be essentially devoid of galaxies. Simulations¹¹⁰ also show large-scale cluster orientation correlations resembling those observed. On the other hand, the rapid evolution of the autocorrelation function requires that the pancakes form at $z \leq 2$, which is uncomfortably recent if smaller structures including galaxies must form subsequently.

The cold DM scenario avoids this latter problem because galaxies and clusters would already have formed by the time of supercluster collapse. Would superclusters and voids arise in a cold DM Universe? There are good reasons to believe that pancakes, filaments, and voids would indeed form, and preliminary indications from N -body simulations suggest that they do^{92,111}. The cold and hot DM fluctuation spectra are identical on the very largest scales, differing only in that the latter is cut off below $\sim 10^{15} M_\odot$ by neutrino free streaming and normalized larger above this scale (see Fig. 2a). The presence of already-formed and partially-virialized substructure in the cold DM case causes supercluster collapse to be less dissipative than in the neutrino picture because the caustics are thickened and because a smaller fraction of the baryons remains as cold, uncondensed gas. Nevertheless, as Dekel¹¹² has shown, persistent flattened or elongated structures can form even in the absence of dissipation as a result of continued expansion in directions orthogonal to the collapse. Indeed, the sharp caustics and highly dissipative shocks in the neutrino picture may produce superclusters that are too flat compared with the observations, even when gravitational interactions with neighbouring superclusters are taken into account.

Another related difference between hot and cold DM is that the hot DM Universe is predicted to have a rather simple cellular structure while the cold DM Universe probably has a considerably richer structure, perhaps more like that observed. In particular, the sizes of superclusters and voids in the cold case should span a fairly broad range. This will be an important test of the models when enough galaxy redshift data become available to indicate the statistics of the void distribution.

What should one expect to find in the voids? In the hot DM picture, all galaxies form from the dense gas along the caustics. Hence, the centres of voids should be entirely empty of galaxies and should contain only low-density DM and hot primordial gas (heated by radiation from pancake shocks and too dilute to have cooled)¹¹³. In the cold DM picture, one might at first suppose that galaxies form more or less uniformly in space, with their density subsequently enhanced by gravitational clustering and the pancake distortion of the Hubble flow. One would then expect to find galaxies in the voids, although with lower density than in superclusters.

In a low-density Universe, both analytical calculations^{98,99} and N -body simulations⁹² suggest that large regions having $\rho/\langle \rho \rangle \leq 0.2$ cannot form by gravitational clustering alone. Nonetheless, large regions with few bright galaxies may actually be more likely in a low density cold DM Universe, for the following reason. On average, large galaxies form late when the average density has dropped well below the critical density. Larger initial fluctuations are thus required to form large galaxies in low-density regions. But larger fluctuations are statistically less likely in low-density regions with the flat, cold DM spectrum. This suppresses formation of large galaxies in moderate-size voids, and formation of clusters in larger voids⁸⁴.

In addition, feedback from nongravitational processes could amplify the number density contrast of bright galaxies compared with the underlying density of dark and baryonic matter. For example, at $z \sim 10$ the average density is highest where pancakes will later occur. This is where most of the earliest galaxies and rich clusters form, and these will be the regions earliest enriched

in metals. Radiation from these early galaxies (and quasars, and possibly Population III stars or VMOs) could heat the gas in lower density regions, raising the Jeans mass, as discussed above, and suppressing small galaxy formation. Indeed, if this early radiation has a hard enough spectrum and heats the gas sufficiently rapidly, it could raise the gas temperature throughout the Universe to 10^6 K and essentially halt galaxy formation outside pancakes. One of us (M.J.R.) has recently discussed¹¹⁴ processes for suppressing galaxy formation in protovoids. Finally, explosive shocks could also enhance galaxy formation in denser regions¹¹⁵. Although the efficacy of all these processes is uncertain and needs further investigation, we conclude that the existence of large regions in which the density of bright galaxies is low is probably not a serious problem for the cold DM picture.

Now let us briefly discuss the two types of correlations mentioned above. The observed correlation of galaxy type with galaxy number density^{103,104} can, at least partly, be understood as a consequence of the greater statistical likelihood of large- σ fluctuations in regions of greater density (protoclusters), together with the effect discussed above that results in higher- σ fluctuations acquiring lower angular momenta on average and becoming elliptical galaxies or spheroidal bulges. There is also an environmental effect: lower- σ fluctuations yield dark haloes that are physically larger and more diffuse than higher- σ fluctuations. Disks, which form from such fluctuations, thus form slowly, by infall of gas from large radii within these extended haloes. Because large haloes have correspondingly large collision cross-sections, few disks can form in regions of high galaxy number density. In dense regions, the halo gas is stripped by collisions and is mixed with enriched gas from galactic winds to become the hot intergalactic medium observed in X rays. It will be interesting to investigate these effects with N -body simulations and a more detailed theory of galaxy formation.

Regarding the second type of correlation, Binggeli¹⁰⁵ has found that the position angles of nearby, elongated Abell clusters are within 45° of the direction to the nearest cluster, provided the two clusters are separated by $\leq 15h^{-1}$ Mpc. He found a similar, though less significant, correlation between the position angle of the brightest cluster galaxy's major axis and the direction to the nearest cluster. In the hot DM picture, where superclusters form before clusters and galaxies, such correlations arise naturally. Indeed an N -body simulation¹¹⁰ has shown that such correlations also occur, but more weakly, even with a fluctuation spectrum having an adiabatic peak at $\sim 10^{15} M_\odot$ superimposed on an $n=0$ spectrum, so that rich clusters and superclusters form almost simultaneously. However, a preliminary analysis¹¹⁶ of cold DM N -body simulations finds Binggeli-type correlations only if $\Omega h \leq 0.2$.

It is important to find other observationally accessible information that can discriminate between cosmological models. The large-scale velocity fields of superclusters should be rather different in the hot and cold DM schemes because of the much greater dissipation in the former. With new instruments it will also be possible to study the $z < 2$ evolution of superclustering of quasars and of Ly α absorbing clouds, and perhaps the density and composition of the gas in voids.

It was once hoped that percolation analysis of the large scale galaxy distribution could help to distinguish between different cosmological models.^{11,117} However, when this analysis is applied to the CfA galaxy data, the results are found to be a sensitive function of the depth of the survey¹¹⁸. Moreover, realistic N -body simulations of isothermal and adiabatic scenarios (that is, with accurate treatment of gravitational interactions on small as well as large scales) have nearly identical percolation properties^{110,119}. Better statistical tests are needed to compare objectively the large-scale distribution of matter in models and observations.

Conclusions

We have shown that a Universe with ~ 10 times as much cold dark matter as baryonic matter provides a remarkably good fit

to the observed Universe. This model predicts roughly the observed mass range of galaxies, the dissipational nature of galaxy collapse, and the observed Faber-Jackson and Tully-Fisher relations. It also gives dissipationless galactic haloes and clusters. In addition, it may also provide natural explanations for galaxy-environment correlations and for the differences in angular momenta between ellipticals and spiral galaxies. Finally, the cold DM picture seems reasonably consistent with the observed large-scale clustering, including superclusters and voids. In short, it seems to be the best model available and merits close scrutiny and testing.

We thank Dick Bond, Jim Bardeen, Marc Davis, Avishai Dekel, George Efstathiou, Mike Fall, Carlos Frenk, Margaret Geller, Jim Gunn, Jim Peebles, Nicola Vittorio, and Simon White for valuable conversations. In addition, J.R.P. acknowledges the hospitality of the Max Planck Institute for Physics and Astrophysics, Munich during the summer 1983. This work was supported by the Department of Energy under contract no. DE-AC03-76SF00515 and by NSF grants AST-82-11551 and PHY-81-15541. This is Lick Observatory Contribution No. 436.

1. Faber, S. M. & Gallagher, J. S. *Astrophys. J.* **17**, 135-187 (1979).
2. Peebles, P. J. E. *Astrophys. J. Lett.* **263**, L1-L5 (1982).
3. Primack, J. R. & Blumenthal, G. R. *Formation and Evolution of Galaxies and Large Structures in the Universe* (eds Audouze, J. & Tran Thanh Van, J.) 163-183 (Reidel, Dordrecht, 1983).
4. Primack, J. R. & Blumenthal, G. R. *4th Workshop on Grand Unification* (eds Weldon, H. A., Langacker, P. & Steinhart, P. J.) 256-288 (Birkhauser, Boston, 1983).
5. Primack, J. R. & Blumenthal, G. R. *Clusters and Groups of Galaxies* (eds Mardirosian, F., Giuricin, G. & Mezzetti, M.) (Reidel, Dordrecht, in the press).
6. Peebles, P. J. E. *Astrophys. J.* **277**, 470-477 (1984).
7. Blumenthal, G. R. & Primack, J. R. Preprint (SLAC-PUB-3388, 1984).
8. Peebles, P. J. E. *The Large Scale Structure of the Universe* (Princeton University Press, 1980).
9. Efstathiou, G. & Silk, J. *Fundam. Cosmic Phys.* **9**, 1-138 (1983).
10. Lyubimov, V. A. *et al. Phys. Lett.* **B94**, 266-268 (1980).
11. Zeldovich, Ya. B., Einasto, J. & Shandarin, S. F. *Nature* **300**, 407-413 (1982).
12. Shandarin, S. F., Doroshkevich, A. G. & Zeldovich, Ya. B. *Sov. Phys. Usp.* **26**, 46-76 (1983).
13. Pagels, H. R. & Primack, J. R. *Phys. Rev. Lett.* **48**, 223-226 (1982).
14. Olive, K. A. & Turner, M. S. *Phys. Rev. D* **25**, 213-216 (1982).
15. Blumenthal, G. R., Pagels, H. & Primack, J. R. *Nature* **299**, 37-38 (1982).
16. Bond, J. R., Szalay, A. S. & Turner, M. S. *Phys. Rev. Lett.* **48**, 1636-1639 (1982).
17. Bond, J. R. & Szalay, A. S. *Astrophys. J.* **274**, 443-468 (1984).
18. Ipser, J. & Sikivie, P. *Phys. Rev. Lett.* **50**, 925-927 (1983).
19. Peccei, R. & Quinn, H. *Phys. Rev. D* **16**, 1791-1797 (1977).
20. Weinberg, S. *Phys. Rev. Lett.* **40**, 223-226 (1978).
21. Wilczek, F. *Phys. Rev. Lett.* **40**, 279-282 (1978).
22. Preskill, J., Wise, M. & Wilczek, F. *Phys. Lett.* **120B**, 127-132 (1983).
23. Abbott, L. & Sikivie, P. *Phys. Lett.* **120B**, 133-136 (1983).
24. Dine, M. & Fischler, W. *Phys. Lett.* **120B**, 137-141 (1983).
25. Fukugita, M., Watamura, S. & Yoshimura, M. *Phys. Rev. Lett.* **48**, 1522-1525 (1982).
26. Sikivie, P. *Phys. Rev. Lett.* **51**, 1415-1417 (1983); erratum **52**, 695 (1984).
27. Ellis, J., Hagelin, J. S., Nanopoulos, D. V., Olive, K. & Srednicki, M. *Nucl. Phys. B* **238**, 453-476 (1984).
28. Stecker, F. W. & Shafi, Q. *Phys. Rev. Lett.* **50**, 928-931 (1983).
29. Freese, K., Price, R. & Schramm, D. N. *Astrophys. J.* **275**, 405-412 (1983).
30. Carr, B. J. *Comments Astrophys.* **7**, 161-173 (1978).
31. Lacey, C. G. *Formation and Evolution of Galaxies and Large Scale Structures in the Universe* (eds Audouze, J. & Tran Thanh Van, J.) 351-360 (Reidel, Dordrecht, 1984).
32. Lin, D. N. C. & Faber, S. M. *Astrophys. J. Lett.* **266**, L21-L25 (1983).
33. Canizares, C. R. *Astrophys. J.* **263**, 508-517 (1982).
34. Witten, E. *Phys. Rev. D* **30**, 272-285 (1984).
35. Carr, B. J., Bond, J. R. & Arnett, W. D. *Astrophys. J.* **277**, 445-469 (1984).
36. Hegyi, D. J. & Olive, K. A. *Phys. Lett.* **126B**, 28-32 (1983).
37. Yang, J., Turner, M. S., Steigman, G., Schramm, D. N. & Olive, K. A. *Astrophys. J.* (in the press).
38. Uson, J. M. & Wilkinson, D. T. *Astrophys. J. Lett.* **277**, L1-L4 (1984).
39. Uson, J. M. & Wilkinson, D. T. *Inner Space/Outer Space* (eds Kolb, E. W. *et al.*) (University of Chicago Press, in the press).
40. Weinberg, S. *Astrophysical Cosmology: Proc. Study Week on Cosmology and Fundamental Physics* (eds Brück, H. A., Coyne, G. V. & Longair, M. S.) 503-528 (Pontifical Scientific Academy, Vatican, 1982).
41. Kolb, E. W. & Turner, M. S. *A. Rev. Nucl. Part. Sci.* **33**, 645-696 (1983).
42. Rees, M. J. *The Very Early Universe* (eds Gibbons, G., Hawking, S. & Siklos, S.) 29-58 (Cambridge University Press, 1983).
43. Frenk, C., White, S. D. M. & Davis, M. *Astrophys. J.* **271**, 417-430 (1983).
44. Dekel, A. & Aarseth, S. *Astrophys. J.* **283** (in the press).
45. Kaiser, N. *Astrophys. J. Lett.* **273**, L17-L20 (1983).
46. Faber, S. M. *Proc. 1st ESO-CERN Symp., Large Scale Structure of the Universe, Cosmology, and Fundamental Physics* (in the press).
47. Bond, J. R., Szalay, A. S. & White, S. D. M. *Nature* **301**, 584-585 (1983).
48. Aaronson, M. *Astrophys. J. Lett.* **266**, L11-L15 (1983).
49. Aaronson, M. & Cook, K. *Bull. Am. astr. Soc.* **15**, 907 (1983).
50. Cook, K., Schechter, P. & Aaronson, M. *Bull. Am. astr. Soc.* **15**, 907 (1983).
51. Faber, S. M. & Lin, D. N. C. *Astrophys. J. Lett.* **266**, L17-L20 (1983).
52. Davis, M., Lecar, M., Pryor, C. & Witten, E. *Astrophys. J.* **250**, 423 (1981).
53. Hut, P. & White, S. D. M. *Nature* **310**, 637-644 (1984).
54. Doroshkevich, A. G. & Khlopov, M. Yu. *Mon. Not. R. astr. Soc.* (in the press).
55. Turner, M. S., Steigman, G. & Krauss, L. M. Preprint (Fermilab, 1984).
56. Primack, J. R. & Roncadelli, M. Preprint (SLAC-PUB-3304, 1984).
57. Shafi, Q. & Stecker, F. W. Preprint (Bartol Research Foundation, 1984).
58. Gelmini, G., Schramm, D. N. & Valle, J. W. F. Preprint (CERN-TH.3865, 1984).
59. Harrison, E. R. *Phys. Rev. D* **1**, 2726-2730 (1970).
60. Peebles, P. J. E. & Yu, J. T. *Astrophys. J.* **162**, 815-836 (1970).
61. Zeldovich, Ya. B. *Mon. Not. R. astr. Soc.* **160**, 1P-3P (1972).

62. Shafi, Q. & Vilenkin, A. *Phys. Rev. Lett.* **52**, 691-694 (1984).
63. Steinhart, P. J. & Turner, M. S. Preprint (Fermilab, 1984).
64. Quinn, H. R. & Gupta, S. *Phys. Rev. D* **29**, 2791-2797 (1984).
65. Zeldovich, Ya. B. *Mon. Not. R. astr. Soc.* **192**, 663-667 (1980).
66. Vilenkin, A. *Phys. Rev. D* **24**, 2082-2089 (1981).
67. Bond, J. R., Efstathiou, G. & Silk, J. *Phys. Rev. Lett.* **45**, 1980-1984 (1980).
68. Doroshkevich, A. G., Zeldovich, Ya. B., Sunyaev, R. A. & Khlopov, M. Yu. *Soviet Astr. Lett.* **6**, 252-256 (1980).
69. Chernin, A. D. *Soviet Astr.* **25**, 14-16 (1981).
70. White, S. D. M. & Rees, M. J. *Mon. Not. R. astr. Soc.* **183**, 341-358 (1978).
71. Rees, M. J. & Ostriker, J. P. *Mon. Not. R. astr. Soc.* **179**, 541-559 (1977).
72. Faber, S. M. *Astrophysical Cosmology: Proc. Study Week on Cosmology and Fundamental Physics* (eds Brück, H. A., Coyne, G. V. & Longair, M. S.) 191-234 (Pontifical Scientific Academy, Vatican, 1982).
73. Silk, J. *Nature* **301**, 574-578 (1983).
74. Forman, W. & Jones, C. *Ann. Rev. Astr. Astrophys.* **20**, 547-585 (1982).
75. Geller, M. J. & Beers, T. C. *Proc. Astr. Soc. Pac.* **94**, 421-439 (1982).
76. Thuan, T. X. & Romanishin, V. *Astrophys. J.* **248**, 439-459 (1981).
77. Cavaliere, A., Santangelo, P., Tarquini, G. & Vittorio, N. *Clusters and Groups of Galaxies* (eds Mardirossian, F., Giuricin, G. & Mezzetti, M.) (Reidel, Dordrecht, in the press).
78. Yoneyama, T. *Publ. Astr. Soc. Jap.* **24**, 87-98 (1972).
79. Peebles, P. J. E. Preprint (Princeton Univ., 1984).
80. Fall, S. M. & Rees, M. J. (in preparation).
81. Fall, S. M. & Rees, M. J. *Mon. Not. R. astr. Soc.* **181**, 37P-42P (1977).
82. Geller, M. & Huchra, J. *Astrophys. J. Suppl.* **52**, 61-87 (1983).
83. Faber, S. M., Blumenthal, G. R. & Primack, J. R. *Astrophys. J. Lett.* (submitted).
84. Blumenthal, G. R., Faber, S. M. & Primack, J. R. (in preparation).
85. Efstathiou, G. & Jones, B. J. T. *Mon. Not. R. astr. Soc.* **186**, 133-144 (1979).
86. Fall, S. M. & Efstathiou, G. *Mon. Not. R. astr. Soc.* **193**, 189-206 (1980).
87. Sandage, A., Freeman, K. C. & Stokes, N. R. *Astrophys. J.* **160**, 831-844 (1970).
88. Efstathiou, G. & Barnes, J. *Formation and Evolution of Galaxies and Large Scale Structures in the Universe* (eds Audouze, J. & Tran Thanh Van, J.) 361-377 (Reidel, Dordrecht, 1984).
89. Gott, J. R. & Thuan, T. X. *Astrophys. J.* **204**, 649-667 (1976).
90. Davis, M. & Peebles, P. J. E. *Astrophys. J.* **267**, 465-482 (1983).
91. Sandage, A. *Proc. 1st ESO-CERN Symp., Large Scale Structure of the Universe, Cosmology, and Fundamental Physics* (in the press).
92. Davis, M., Efstathiou, G., Frenk, C. & White, S. D. M. (in preparation).
93. Bond, J. R. & Efstathiou, G. Preprint (NSF-ITP-84-79, 1984).
94. Vittorio, N. & Silk, J. Preprint (U. Cal. Berkeley, 1984).
95. Peebles, P. J. E. Preprint (Princeton Univ., 1984).
96. Kaiser, N. *Inner Space/Outer Space* (eds Kolb, E. W. et al.) (University of Chicago Press, in the press).
97. Bardeen, J. *Inner Space/Outer Space* (eds Kolb, E. W. et al.) (University of Chicago Press, in the press).
98. Hoffman, G. L., Salpeter, E. E. & Wasserman, I. *Astrophys. J.* **268**, 527-539 (1983).
99. Hoffman, Y. & Shaham, J. *Astrophys. J. Lett.* **262**, L23-L26 (1982).
100. Kirshner, R. F., Oemler, A., Schechter, P. L. & Schectman, S. A. *Astrophys. J. Lett.* **248**, L57-L60 (1981).
101. Kirshner, R. F., Oemler, A., Schechter, P. L. & Schectman, S. A. *Early Evolution of the Universe and its Present Structure, IAU Symp. No. 104* 197-202 (1983).
102. Oort, J. H. A. *Rev. Astr. Astrophys.* **21**, 373-428 (1983).
103. Dressler, A. *Astrophys. J.* **236**, 351-365 (1980).
104. Postman, M. & Geller, M. J. *Astrophys. J.* **281**, 95-99 (1984).
105. Binggeli, B. *Astr. Astrophys.* **107**, 338-349 (1982).
106. Zeldovich, Ya. B. *Astr. Astrophys. S.* **84**, 84-89 (1970).
107. Arnold, V. I., Shandarin, S. F. & Zeldovich, Ya. B. *Geophys. Astrophys. Fluid Dyn.* **20**, 111 (1982).
108. Bond, J. R., Centrella, J., Szalay, A. S. & Wilson, J. R. *Formation and Evolution of Galaxies and Large Structures in the Universe* (eds Audouze, J. & Tran Thanh Van, J.) 87-99 (Reidel, Dordrecht, 1984).
109. Shapiro, P. R., Struck-Marcell, C. & Melott, A. L. *Astrophys. J.* **275**, 413-429 (1983).
110. Dekel, A., West, M. J. & Aarseth, S. J. *Astrophys. J.* **279**, 1 (1984).
111. Melott, A., Einasto, J., Saar, E., Suisalu, I., Klypin, A. A. & Shandarin, S. F. *Phys. Rev. Lett.* **51**, 935-938 (1983).
112. Dekel, A. *Astrophys. J.* **264**, 373-391 (1983).
113. Doroshkevich, A. G., Shandarin, S. F. & Zeldovich, Ya. B. *Comm. Astrophys.* **9**, 265-273 (1982).
114. Rees, M. J. *Clusters and Groups of Galaxies* (eds Mardirossian, F., Giuricin, G. & Mezzetti, M.) (Reidel, Dordrecht, in the press).
115. Ostriker, J. P. *Astrophysical Cosmology: Proc. Study Week on Cosmology and Fundamental Physics* (eds Brück, H. A., Coyne, G. V. & Longair, M. S.) 473-493 (Pontifical Scientific Academy, Vatican, 1982).
116. Dekel, A. *8th Johns Hopkins Workshop on Current Problems in Particle Theory* (eds Domokos, G. & Kovesi-Domokos, S.) (World Scientific Publishing, Singapore, in the press).
117. Shandarin, S. F. *Soviet Astr. Lett.* **9**, 104 (1983).
118. Barrow, J. D. & Bhavsar, S. P. *Mon. Not. R. astr. Soc.* **205**, 66 (1983).
119. Dekel, A. & West, M. J. Preprint (NSF-ITP-84-44, 1984).
120. Dressler, A. *Astrophys. J.* **226**, 55-69 (1978).
121. Jones, C. & Forman, W. *Astrophys. J.* **276**, 38-55 (1983).
122. Kriss, G. A., Cioffi, D. F. & Canizares, C. R. *Astrophys. J.* **272**, 439-448 (1983).
123. Beers, T. C., Geller, M. J., Huchra, J. P., Latham, D. W. & Davis, R. J. *Astrophys. J.* (in the press).
124. Gunn, J. E. *Astrophysical Cosmology* (eds Brück, H. A., Coyne, G. V. & Longair, M. S.) 233-259 (Pontifical Scientific Academy, Vatican, 1982).
125. Young, P. J. *Astr. J.* **81**, 807-816 (1976).
126. Kent, S. M. & Gunn, J. E. *Astr. J.* **87**, 945-971 (1982).
127. Kent, S. M. & Sargent, W. L. W. *Astr. J.* **88**, 697-708 (1983).
128. de Vaucouleurs, G., de Vaucouleurs, A. & Corwin, H. R. *2nd Reference Catalogue of Bright Galaxies* (University of Texas, Austin, 1976).
129. Dressler, A. & Sandage, A. R. *Astrophys. J.* **265**, 664-680 (1983).
130. Rubin, V. C. *Internal Kinematics and Dynamics of Galaxies* (ed. Athanassoula, E.) 3-6 (Reidel, Dordrecht, 1983).
131. Rubin, V. C., Ford, W. K., Thonnard, N. & Burstein, D. *Astrophys. J.* **261**, 439-456 (1982).
132. Burstein, D., Rubin, V. C., Thonnard, N. & Ford, W. K. *Astrophys. J.* **253**, 70-85 (1982).
133. Thuan, T. X. & Seitzer, P. O. *Astrophys. J.* **231**, 680-687 (1979).
134. Tonry, J. & Davis, M. *Astrophys. J.* **246**, 680-695 (1981).
135. Raymond, J. C., Cox, D. P. & Smith, B. W. *Astrophys. J.* **204**, 290-292 (1976).
136. Dalgarno, A. & McCray, R. A. A. *Rev. Astr. Astrophys.* **10**, 375-426 (1972).

ARTICLES

Ultrahigh gradient particle acceleration by intense laser-driven plasma density waves

C. Joshi*, W. B. Mori*, T. Katsouleas*, J. M. Dawson*,
J. M. Kindel† & D. W. Forslund†

* University of California Los Angeles, California 90024, USA

† Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

Space-charge waves driven by resonantly beating two laser beams in a high-density plasma can produce ultrahigh electric fields that propagate with velocities close to c. By phase-locking particles in such a wave, particles may be accelerated to very high energies within a very short distance.

DURING the past four decades, we have witnessed an increase of six orders of magnitude in the output energy of high-energy accelerators, while the cost per MeV has been reduced by a factor of 16 per decade. But can this progress continue? Current accelerators, such as the Stanford linac, have accelerating fields of 200 keV cm⁻¹. However, for particle energies beyond 10 TeV, one had to invent schemes that can produce fields of at least 10 MeV cm⁻¹. In any particle accelerator scheme, the basic requirement for obtaining particles with ultrahigh energies is an intense longitudinal electric field that interacts with particles for a long time. Since highly relativistic particles move nearly at the speed of light *c*, the energy gained by the particles, $\int \mathbf{E} \cdot d\mathbf{l}$, is maximum if the field is made to propagate with the particles. Extremely large electric fields propagating with phase velocities close to *c* can be produced by space charge waves in a plasma (ionized gas). The maximum electric field that can be produced

by such a wave is approximately $\sqrt{n_e} \text{ V cm}^{-1}$, where n_e is the plasma electron density per cm³. Thus for plasma densities in the range 10¹⁶-10²⁰ electrons cm⁻³, the longitudinal electric fields E_L can be as large as 10⁸-10¹⁰ V cm⁻¹. We now show that such high-gradient, high-phase velocity plasma density waves can be driven by intense laser beams. If particles could be phase-locked in such waves, this scheme has the potential for accelerating particles to ultrahigh energies in very short distances.

Theory

If an intense laser beam is propagated in a plasma, then in certain conditions, the transverse electric field of the laser (which may reach values of 10⁹-10¹⁰ V cm⁻¹) can be very effectively transformed into a longitudinal electric field of a plasma density wave. In the laser accelerator scheme known as the 'Plasma beat

wave accelerator' proposed by Tajima and Dawson^{1,7}, such a plasma wave can be driven by beating two colinear laser beams, with frequencies and wavenumbers (ω_0, k_0) and (ω_1, k_1) , in a plasma resonantly, such that the frequency and the wavenumber of the plasma wave are

$$\begin{aligned}\omega_{\text{epw}} &= \omega_0 - \omega_1 \\ k_{\text{epw}} &= k_0 - k_1\end{aligned}\quad (1)$$

To achieve the ultrahigh accelerating gradients, high-density plasmas ($10^{16} < n_e < 10^{20} \text{ cm}^{-3}$) must be used. Also, because the laser of frequency ω_0 cannot penetrate a plasma whose density exceeds the critical density n_c , corresponding to the plasma frequency $\omega_p = \omega_0$, plasmas with densities less than the quarter critical density must be used. Physically, the plasma wave consists of regions of space charge, which propagate with a phase velocity v_{ph} that is equal to the group velocity of the beat wave $v_g = c(1 - \omega_p^2/\omega_0^2)^{1/2}$. These arise because the spatial intensity gradient of the beat wave envelope, which is in the direction of propagation, exerts a periodic force (ponderomotive force) on the plasma at wavenumber k_{epw} . The plasma wave is thus an electrostatic wave with $\mathbf{E}_{\text{epw}}$ parallel to $\mathbf{k}_{\text{epw}}$.

When two parallel propagating laser beams beat in a plasma resonantly, the plasma density fluctuations grow rapidly. Using fluid equations it can be shown that for $v_0/c \ll 1$, the plasma wave electric field, which is proportional to the perturbed density, initially grows in time with a growth rate

$$\delta = \left(\frac{1}{4} \frac{v_0(0)}{c} \frac{v_0(1)}{c} \omega_p \right) s^{-1} \quad (2)$$

where $[v_0(0, 1)]/c = eE_{(0, 1)}/m\omega_{(0, 1)}c$ is the normalized oscillatory velocity of an electron in the laser field. Wavebreaking is approached when the perturbed density becomes as large as the initial density; however, to reach this limit the plasma wave must remain in phase with the beat wave. As the plasma wave grows, the relativistic effect on the frequency mismatch becomes important² and the plasma wave saturates at a lower amplitude given by

$$\varepsilon = \frac{eE_{\text{epw}}}{m\omega_{\text{epw}}c} = \left(\frac{16}{3} \frac{v_0(0)}{c} \frac{v_0(1)}{c} \right)^{1/3} \quad (3)$$

Once the saturation amplitude is reached, the plasma wave amplitude actually begins to decrease, if this process alone acts to sustain the plasma wave³.

As the plasma wave is an electrostatic wave, we can use Poisson's equation, $E_{\text{epw}} = 4\pi n e n_e / k_{\text{epw}}$, to estimate the maximum electric field that we might expect if we assume that the perturbed electron density is equal to the initial density (the so-called cold plasma wavebreaking limit $\varepsilon = 1$). As the plasma wave is propagating at relativistic speed, the Lorentz transformation gives the maximum energy gained by an electron with initial energy $\gamma_{\text{ph}} mc^2$ as $2\gamma_{\text{ph}}^2 mc^2$ or $(n_e/n_c) \text{ MeV}$ in a distance $2\gamma_{\text{ph}}^2 c/\omega_p$. Here, $\gamma_{\text{ph}} = (\omega_0/\omega_p)$. For a given laser wavelength, the maximum energy gained increases as the plasma density is reduced but because the electric field scales as $\sqrt{n_e}$ it takes longer and longer to obtain this energy.

In a practical accelerator, we cannot use wavebreaking electric fields because as the electric field approaches this value it 'traps' background plasma electrons and begins to accelerate them. It also develops harmonics. Both of these effects are not desirable if we wish to accelerate an externally-injected bunch of electrons. Computer simulations (described below) show that a coherent plasma wave, without any significant number of self-trapped particles, can be generated with electric fields of the order of 10% of the wavebreaking electric field. If we require that the minimum value of the longitudinal field be 10 MeV cm^{-1} , this leads to a minimum plasma density of 10^{16} cm^{-3} . Using a CO_2 laser to excite a beat wave in such a plasma, gives maximum particle energies of $\sim 100 \text{ MeV}$. The physical reason for this limit on the maximum energy is the dephasing of particles and the electric field of the correct polarity (which only exists for half-wavelength of the plasma wave in the wave frame). Although

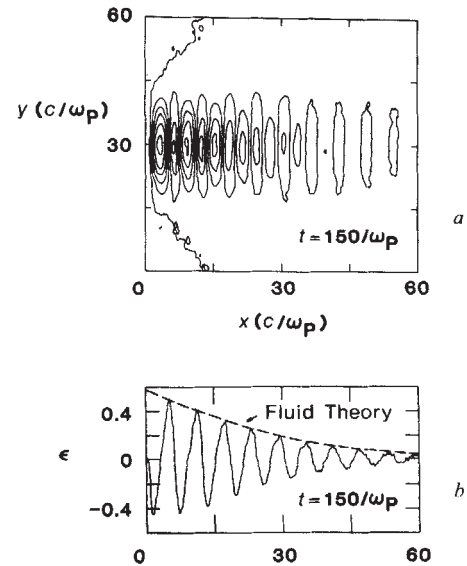


Fig. 1 Potential contours of the space-charge density wave driven by the two laser beams at time $150/\omega_p$ (a) section through the centre of the longitudinal electric field of the plasma wave (b). The dotted line in b shows a comparison of the rate of plasma wave build-up as predicted by the fluid theory with that obtained in these two-dimensional simulations.

for a given plasma density the maximum energy scales inversely with square of the laser wavelength λ , it becomes increasingly difficult to keep the laser beam focused at shorter wavelengths because the Rayleigh length scales as λ . In a single stage, therefore, the best that can probably be achieved is to accelerate electrons to a few GeV using a laser with wavelength in the micrometre range.

To obtain higher energies, one has to be prepared for multi-staging or invent a scheme for phase-locking the particles in the wave so that they do not out run the wave. One such scheme will be described later.

Simulations

Computer simulations were carried out using the two-dimensional fully relativistic, electromagnetic particle code WAVE⁴ to study the above described acceleration mechanism in a self-consistent manner. A typical WAVE simulation has 10^6 electrons and 10^6 ions. From the particle positions and velocities, charge densities and current densities are found. Then Maxwell's equations are used to calculate the electric and magnetic fields or equivalently the forces on the particles. Equations of motion are used, to calculate new particle locations and velocities and then the process is repeated.

The results of a typical two-dimensional simulation are depicted in Figs 1-3. The parameters for this simulation were as follows: two laser beams $\omega_0 = 5\omega_p$ and $\omega_1 = 4\omega_p$, each with an r.m.s. intensity of $v_0 = 0.4 c$, a transverse intensity profile of $\cos^2 y$ and a beam width of $30 c/\omega_p$ were injected into a homogeneous plasma $60 c/\omega_p$ long and $60 c/\omega_p$ wide with a temperature of 2.5 keV . The laser beams had a cubic rise from zero to maximum intensity in $300/\omega_p$ after which the laser intensity remained constant. The ion-to-electron mass ratio was 1,836 and temperature ratio was 1. Thus for an IR laser with wavelengths $9.6 \mu\text{m}$ and $12 \mu\text{m}$ the parameters were: a hydrogen plasma with $n_e \sim 4.9 \times 10^{17} \text{ cm}^{-3}$, laser risetime $\sim 7.6 \text{ ps}$, laser intensity $\sim 2 \times 10^{15} \text{ W cm}^{-2}$, beam width ~ 25 wavelengths and system length ~ 50 wavelengths. The simulation parameters were chosen to study both the beat wave driven plasma wave and other competing one- and two-dimensional effects.

The simulation results can be categorized into two time regimes. The first, during the risetime of the laser pulse when

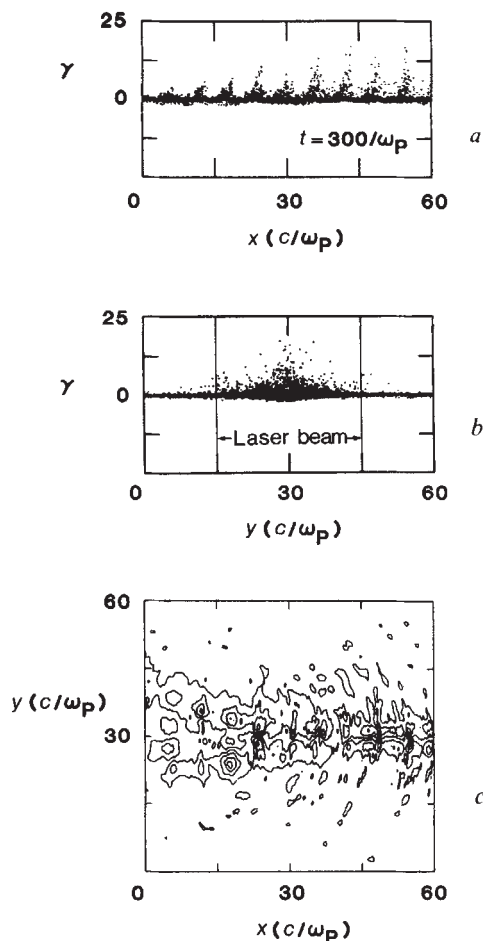


Fig. 2 Phase space plot (momentum in the x direction) of the plasma electrons shows that at time $300/\omega_p$, trapped electrons are accelerated in every wavelength of the plasma wave which represents an accelerating 'bucket' (a). Particles have larger energies on the right-hand side as they have seen an accelerating field for a longer time. The maximum particle energies of $\gamma = 25$ are in good agreement with theory. In the transverse direction, although the laser beam is $30 c/\omega_p$ wide, most particles are confined to a narrow region down the centre of the accelerating wave (b). Contours of the self-consistent magnetic field at time $t = 300/\omega_p$ generated by the accelerating trapped particles (c). This field exerts a radial force on the trapped particles confining them in a small region down the axis and also forces the return current to flow on the outside of the region where it exists.

a very coherent plasma wave grows and reaches saturation and the second after the peak of the laser pulse ($300/\omega_p$) when competing phenomena come into play.

Figure 1a shows the contour plot of the potential of the plasma density wave at $150/\omega_p$. According to the fluid theory, the plasma wave should reach peak amplitude at this time. The plasma wave wavefronts in the contour plot are seen to be planar. A section through the centre of the transverse axis of the longitudinal electric field plot is shown in Fig. 1b. An extremely coherent plasma wave builds up rapidly in time/space and saturates at $\epsilon \approx 0.5$, in excellent agreement with the value expected from the fluid theory. By $300/\omega_p$, the longitudinal electric field becomes non-planar as the laser beams begin to self-focus. Fluid theory would predict that once saturation is reached, the plasma wave amplitude decreases but simulations show that a 'plasma wave' with greatly reduced coherence persists driven by the multiple Raman forward scattering.

Figure 2a shows that the electrons trapped at an earlier time, when the plasma wave field was intense and coherent, have by time $300/\omega_p$ approached maximum energies. Accelerated parti-

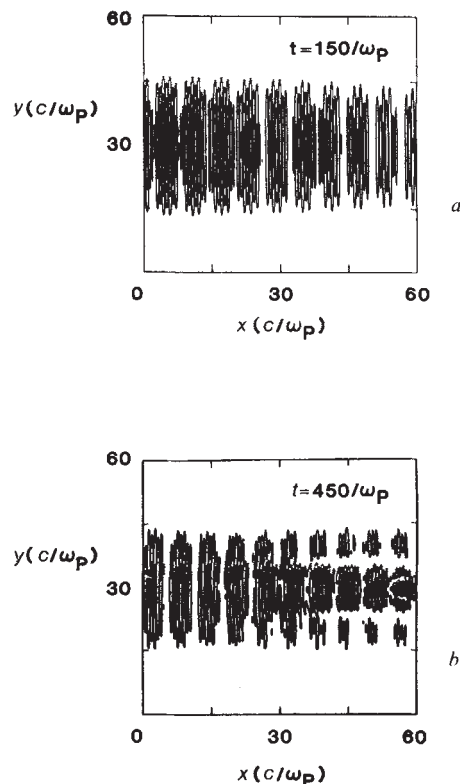


Fig. 3 Contour plot of the beat pattern produced by two colinearly-propagating laser beams in the plasma at time $150/\omega_p$ (a) and at $450/\omega_p$ (b). The x and y dimensions are in units of c/ω_p . Clear evidence for laser beam self-focusing is seen at the later time from the converging inner contours. As a result of self-focusing, the plasma is pushed out and the space-charge wave accelerating the particles disrupts.

cles leaving the right-hand boundary of the plasma (not shown here) have maximum energies of $\gamma = 25$, in reasonable agreement with theory. Here $\gamma = (1 - \beta^2)^{-1/2}$ is the usual relativistic scaling factor. More important, however, is the fact that although the laser beam and, therefore, the plasma wave is $30 c/\omega_p$ wide, most of the accelerated particles are collimated within $10 c/\omega_p$ of the centre (Fig. 2b). The accelerating particles remain highly collimated because the trapped particles give rise to an azimuthal magnetic field which, in turn, exerts a radial force on the particles keeping them tightly focused. The contour plot of this self-generated magnetic field which can reach megagauss magnitudes is shown in Fig. 2c. The maximum field is in the regions where bunches of electrons are accelerated. The accelerated particles constitute a flow of charges or current which leaves the plasma positively charged, thus driving a return current of colder particles in the opposite direction. In the simulations the return current is found to flow predominantly on the outside of the region bounded by the self-generated magnetic field.

These two-dimensional simulations also allowed us to investigate competition between the beat wave driven plasma wave and Raman back- and side-scattering. To allow the growth of large angle side-scattering the transverse dimension of the beam was made periodic. Even in relatively cold plasmas ($T_e \sim 200$ eV), where Landau damping is expected to be weak, Raman backscatter is not found to be a problem under two-frequency illumination. Small angle (from the forward direction) side-scatter can occur, effectively making the desired plasma wave non-planar. This gives rise to a radial component of the electric field; however, the self-generated magnetic field tends to keep the trapped particles focused in the centre of the plasma wave as discussed above.

Figure 3a shows the laser beat wave contours at $150/\omega_p$ as the beams propagate from left to right. Contrast this with laser

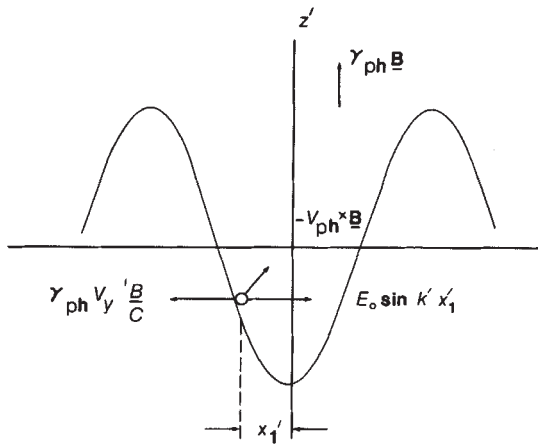


Fig. 4 In the phase stabilization scheme, Surfatron, a perpendicular magnetic field is applied to the wavevector of the longitudinal space-charge wave. In the wave frame, a trapped particle comes to an equilibrium position against the side of the potential well where the 'trapping' and 'de-trapping' forces balance: $E_0 \sin k' x_1' = \gamma_{ph} B$. The energy gained by the particle in the direction of the wave is simply $\gamma_{ph} B x_1'$.

beam contours at $450/\omega_p$ (Fig. 3b). Although the outer contours are still the same size, the contours at the beam centre are seen to 'self-focus' to a smaller size. We have determined that this self-focusing is due to two effects: intensity-dependent refractive index (relativistic self-focusing) and transverse intensity gradient (ponderomotive self-focusing). Whether the beam self-focuses as a whole or breaks up into smaller filaments depends on the transverse size and the intensity profile of the laser beams. In any case, catastrophic self-focusing is an undesirable phenomenon because it tends to push the plasma out and create a channel of low density, thereby destroying the resonance condition for driving the plasma wave and terminating the accelerating process.

Two-dimensional simulations of the beat wave build-up have been encouraging and given results that are consistent with predictions of the fluid and single particle theories. Although there are many competing instabilities in the plasma, these occur on a longer time scale; and plasma wave coherence can be maintained over typically tens of plasma wave wavelengths at the front of the laser pulse, which continually moves into unperturbed plasma. If the laser pulse is longer than this, then wave-particle and wave-wave interactions heat up the plasma to temperatures of hundreds of keV at the back of the laser pulse and make it extremely turbulent. Thus, extremely short laser pulses on the order of a few picoseconds are probably most effective in this acceleration scheme. Such short pulses will also inhibit other instabilities such as the stimulated Brillouin scattering.

Phase-stabilization

The limit on energy gain of a particle in the Plasma beat wave accelerator can be overcome if a magnetic field of appropriate strength, $\gamma_{ph} B < E_{epw}$, is applied perpendicular to the plasma wave (Fig. 4). This accelerates particles parallel to the phase fronts of the accelerating wave, which, in turn, provides a force keeping the particles in phase with the wave⁵. In principle, therefore, particles can gain energy at rate

$$\frac{\Delta U}{\Delta x} = 0.1 \left[\frac{B_{KG}}{n_{16} \lambda_\mu} \right] (n_{16})^{1/2} \text{ GeV cm}^{-1} \quad (4)$$

in the direction of the wave and the accelerated particles move at an angle $\tan \theta = 3 \times 10^{-3} \sqrt{n_{16} \lambda_\mu}$ with respect to the wavevector of the plasma wave. Here B_{KG} is the applied magnetic field in kilogauss, n_{16} is the plasma density in units of 10^{16} cm^{-3} and λ_μ is the laser wavelength in micrometres. The quantity in the

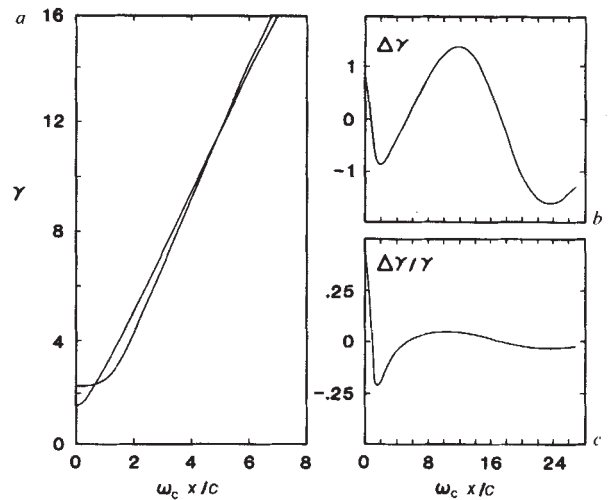


Fig. 5 *a*, Energy gain (γ) plotted against distance for two Surfatron particles of different initial velocities (0.8c and 0.9c). In *b*, energy separation ($\Delta\gamma$) of the particles remains roughly constant, while in *c*, $\Delta\gamma/\gamma$ decreases rapidly. $V_{ph}/c = .9$, $cE/B = 2.5$, $\omega/\omega_c = 9$. ω_c is the nonrelativistic electron cyclotron frequency.

square bracket must be $< \varepsilon$ to prevent the detrapping by the magnetic field. Because the particles move at an angle with respect to the plasma wave, the width of the plasma wave necessary to accelerate the particles to the desired energy can be rather large. Fortunately, the plasma wave width (and hence the energy required to excite it) can be reduced considerably by beating the electromagnetic waves at a finite angle (optical mixing) rather than colinearly⁶. One possible limitation associated with finite angle optical mixing is that, while the particles move at $\sim c$, the laser energy propagates through the plasma with the particles at v_g . In a device of length L the particles move ahead with respect to the laser pulse by a distance $\Delta L = \omega_p^2 L / \omega_0^2$. Thus the plasma wave coherence must be maintained over a distance ΔL behind the leading edge of the laser pulse. We have carried out one-dimensional (one spatial, three velocities) particle simulations, which confirm the above scaling laws. Two-dimensional computer simulations are being carried out to test the phase-locking and cross-field acceleration concept when the plasma wave is driven by finite angle optical mixing, but this technique has the promise of utilizing the high gradients of space-charge plasma waves for obtaining ultrahigh energies. For example, using a 100-kG magnetic field across a 10^{19} cm^{-3} plasma and a 1- μm laser, particles may be accelerated to 100 GeV in a device that is only 3.16 m long.

Laser and plasma requirements

While there is a tremendous promise for $> 10 \text{ MeV cm}^{-1}$ acceleration gradients in this scheme, there are technological obstacles between the concept and a realizable working system. The laser needed for a TeV class accelerator based on this concept will exceed the capability of fusion lasers (presently the largest laser systems) in the areas of peak power, pulse width, repetition rate, and the need for the power to be delivered in a single diffraction limited beam. Considerable work is also needed in the areas of focusing and optical transport of such high peak power laser beams.

An estimate of the laser power required can be obtained from energy balance arguments. The longitudinal electric field of the plasma wave represents an energy density

$$W_{epw} = \frac{E_{epw}^2}{8\pi} \approx 500 n_{16} \varepsilon^2 \text{ J cm}^{-3} \quad (5)$$

Because the group velocity of the plasma wave, $3v_e^2/v_{ph} \ll c$, we assume that the laser pulse excites the plasma wave only in the

volume through which it propagates. To avoid significant pump depletion, the total energy in the laser beams must be much greater than that in the plasma wave.

The problem of producing plasmas (particularly with a cross magnetic field) with densities between $10^{16} < n_e < 10^{20} \text{ cm}^{-3}$ and with the required length and homogeneity will also require considerable attention. At high intensities that are necessary in this scheme, the requirement that plasma frequency be exactly equal to the difference frequency may be relaxed, but this issue needs careful investigation.

Prospects for ultrahigh energies

High-phase velocity space charge waves in a plasma have the potential for producing the high accelerating gradients that are necessary for a new generation of particle accelerators. Whether this scheme can be used to accelerate particles to ultrahigh energies depends on how well the phase stabilization (Surfatron) scheme can be made to work in practice. Particular problems are pump depletion, filamentation of the laser and injected particle beams, and the effect of self-generated magnetic fields on the accelerating particles.

On the positive side, the Surfatron scheme would produce a compact accelerator with a high quality beam. The energy spread, $\Delta\gamma/\gamma$ would be small because approximately half of the injected particles quickly form accelerating 'buckets' that are narrow in phase space near the equilibrium point. In the wave frame this is where $\gamma_{ph}B$ is equal to the local longitudinal electric field. Thus all the particles gain energy at nearly the

same rate, the one given by equation (4). The bunch length of the final beam as well as $\Delta\gamma/\gamma$ is thus expected to be extremely small, as shown in Fig. 5.

The number of accelerated particles N can be estimated from the consideration of the energy balance⁶:

$$N = 2 \times 10^{10} \frac{\text{laser energy (kJ)}}{\text{particle energy (GeV)}} \varepsilon \alpha \lambda_\mu \sqrt{n_{16}} \quad (6)$$

where factor α is the ratio of particle energy to wave energy.

The zero-order motion of the particles in the Surfatron scheme is approximately a straight line, and thus synchrotron radiation loss turns out not to be a problem. The radiation loss due to high-order bounce motion⁵ and due to oscillation in the laser field, as well as other loss mechanisms such as Coulomb scattering and bremsstrahlung radiation, are also not thought to be serious problems.

Finally, the acceleration scheme is still at an embryonic stage for a reliable estimate to be made of its potential total efficiency from wall power (ac) to beam power. The severest limitation, one that is common to all laser acceleration schemes, is the a.c.-to-laser beam efficiency. If we optimistically assume this to be 10% then using our simulations as a guide we may be able to achieve a total efficiency between 10^{-3} and 10^{-4} .

We thank Professor F. F. Chen and Dr C. E. Clayton for valuable comments and discussions. This work was supported at UCLA by DOE contract DE-AM03-76SF00034, NSF grant ECS 83-10972 and LLNL University Research Program and at LANL by the DOE.

Received 11 May; accepted 6 August 1984.

1. Tajima T. & Dawson, J. M. *Phys. Rev. Lett.* **43**, 267-270 (1979).
2. Rosenbluth, M. N. & Liu, C. S. *Phys. Rev. Lett.* **29**, 701-705 (1972).
3. Joshi, C., Tajima, T., & Dawson, J. M., Baldi, H. A. & Ebrahim, N. A. *Phys. Rev. Lett.* **47**, 1285-1288 (1981).

4. Morse, R. L. & Neilson, C. W. *Phys. Fluids* **14**, 830-840 (1971).
5. Katsouleas, T. & Dawson, J. M. *Phys. Rev. Lett.* **51**, 392 (1983).
6. Katsouleas, T., Joshi, C., Mori, W. B. & Dawson, J. M. *Proc. 12th int. Conf. on High-Energy Accelerators* (Fermilab, Batavia, 1982).
7. Cohen, E. I., Kaufman, A. N. & Watson, K. M. *Phys. Rev. Lett.* **29**, 581-584 (1972).

Is there a climatic attractor?

C. Nicolis

Institut d'Aéronomie Spatiale de Belgique, 3 avenue Circulaire, 1180 Bruxelles, Belgium

G. Nicolis

Faculté des Sciences de l'Université Libre de Bruxelles, Campus Plaine, Boulevard du Triomphe, 1050 Bruxelles, Belgium

Much of our information on climatic evolution during the past million years comes from the time series describing the isotope record of deep-sea cores. A major task of climatology is to identify, from this apparently limited amount of information, the essential features of climate viewed as a dynamic system. Using the theory of nonlinear dynamic systems we show how certain key properties of climate can be determined solely from time series data.

EXPERIMENTAL data have essentially two roles in the process of modelling. First, they parameterize the equations postulated by the modeller. Second, they set constraints to be satisfied by the model. For instance, a reasonable model of Quaternary glaciations should reproduce the general aspects of a palaeo-temperature time series as deduced from ice or deep-sea core data and, in particular, should exhibit the characteristic time scales of 100,000, 41,000 and 22,000 yr (ref. 1). In either case, however, the information drawn from the data will remain essentially one-dimensional. Thus, starting from the time series of a certain variable one may construct a power spectrum or a histogram which, despite their interest, do not provide any hint about the additional variables that may affect the evolution. We show here that experimental data contain far richer information which, independent of any particular model, can be used to 'resurrect' the multivariable dynamics of a system starting from a time series pertaining to a single variable.

We should first comment on the status of a time series from the standpoint of the theory of dynamical systems. Let $X_0(t)$ be the time series available from the data, and $\{X_k(t)\}$, where

$k = 0, 1, \dots, n-1$, the full set of variables actually taking part in the dynamics. $\{X_k\}$ is expected to satisfy a set of first-order nonlinear equations, whose form is generally unknown but which, given a set of initial data $\{X_k(0)\}$, will produce the full details of the system's evolution. It is instructive to visualize this evolution in an abstract multi-dimensional space spanned by these variables, the phase space. An instantaneous state of the system becomes a point, say P, in this space, whereas a sequence of such states followed as time varies defines a curve, the phase space trajectory (see Fig. 1). As time grows and transients die out, the system is expected to reach a state of permanent regime, not necessarily time-independent. In phase space, this will be reflected by the convergence of whole families of phase trajectories towards a subset of phase space (C in Fig. 1), such that the system subsequently remains trapped therein. We refer to this invariant set as the attractor.

The interest of the phase space description of a system lies primarily in the fact that the nature of the attractors provides extensive information on the time behaviour of the variables and on the nature of their coupling. For instance, a point

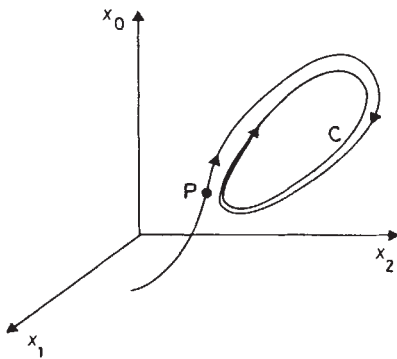


Fig. 1 Typical phase space trajectory emanating from point P and converging to a periodic attractor, represented by curve C.

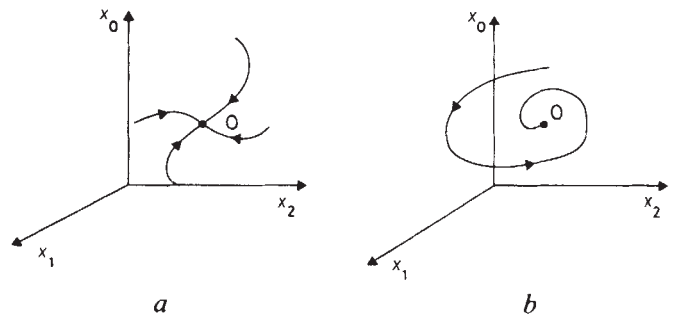


Fig. 2 Two different approaches to a steady state, represented by point 0 in phase space (point attractor). *a*, Monotonic approach; *b*, approach through damped oscillations.

attractor (point 0 in Fig. 2*a, b*) implies that the transient-free behaviour of the system will be time-independent. In addition, if the attractor is approached in the way shown in Fig. 2*a* the transients will die out monotonically, whereas a spiral approach like that in Fig. 2*b* indicates that transients perform damped oscillations. An attractor in the form of a closed curve (C in Fig. 1) implies, on the other hand, that the system will perform sustained oscillations in time with an intrinsically-determined period and amplitude. Finally, multiperiodic motions with incommensurate frequencies are represented in phase space by high-dimensional attracting toroidal surfaces. In each of the above cases, the character of the attractor also gives an indication of the minimum number of variables that should be involved in the description. Thus, for the zero-dimensional attractors of Fig. 2 ($d=0$) we need, respectively, at least one variable ($n=1$) for Fig. 2*a* and two variables ($n=2$) for Fig. 2*b*. For the one-dimensional attractor of Fig. 1 ($d=1$) we need at least two variables ($n=2$), and for the simplest quasi-periodic phenomenon we need at least a two-dimensional torus ($d=2$) embedded in a three-dimensional phase space ($n=3$).

So far we have argued in terms of attractors which are points, lines or surfaces or, in more technical terms, smooth topological manifolds characterized by an integer dimensionality. In recent years, it has been firmly established that geometrical objects exist which are not topological manifolds. Such constructions, which have a non-integer dimensionality, are known as fractals². The theory of dynamical systems and some experimental results from fluid dynamics and chemical kinetics provide us with many examples of fractal attractors³. The importance of the latter stems from the fact that they model irregular, time-dependent phenomena, characterized by two features, both of which are shared by the climatic system: a marked sensitivity to initial conditions; and the appearance of large dispersions from a mean motion similar to a stochastic process, even though the underlying dynamics is perfectly deterministic. Note that the dimensionality of an attractor, fractal or smooth, is bound to be a number smaller than the number of variables present in the evolution.

Climatic attractor

We shall establish the existence of an attractor associated with the long-term climatic evolution of the past million years and determine its dimensionality. Knowing the characteristics of this climatic attractor automatically gives us information on the minimum number of variables that must be introduced into the description.

Our starting point is the oxygen isotope record obtained from the equatorial Pacific deep-sea core V28-238 (refs 4, 5), which is one of the best climatic records available, because the sediment accumulated at a fairly regular rate⁶.

Let $X_0(t)$ be the corresponding time series⁷. Because the n variables $\{X_k(t)\}$ satisfy a set of first-order differential equations, successive differentiation in time reduces the problem to a single (generally highly nonlinear) differential equation of n th order

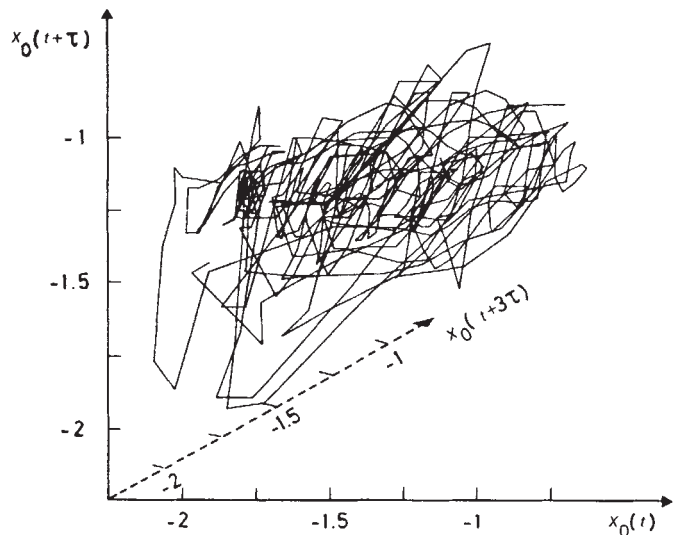


Fig. 3 A view of the climatic attractor embedded in a three-dimensional phase space. The result utilizes about 500 equidistant values of X_0 . These are inferred from the oxygen isotope record obtained from the V28-238 deep sea core of Shackleton and Opdyke⁴ and extended over the past million years, following an interpolation available from the data bank of the University of Louvain⁷. The value of the shift τ adopted in drawing the figure is $\tau = 2$ kyr.

for one of these variables. Thus, instead of $X_k(t)$, $k=0, 1, \dots, n-1$, we may consider $X_0(t)$, the variable supplied by the time series data, and its $n-1$ successive derivatives $X_0^{(k)}(t)$, $k=1, \dots, n-1$, to be the n variables of the problem spanning the phase space of the system⁸. Now, both X_0 and its derivatives can be deduced from the single time series pertaining to $X_0(t)$, as provided by the data. We see, therefore, that, in principle we have sufficient information at our disposal to go beyond the one-dimensional space of the original time series and to unfold the system's dynamics into a multidimensional phase space.

Actually, as suggested originally by Ruelle⁹, instead of $X_0(t)$ and $X_0^{(k)}(t)$ it will be easier to work with $X_0(t)$, and the set of variables obtained from it, by shifting its values by a fixed lag τ . We, therefore, consider from now on the phase space defined by the variables:

$$X_0(t), X_0(t+\tau), \dots, X_0(t+(n-1)\tau) \quad (1)$$

For a typical choice of τ these variables are expected to be linearly independent, and this is all one needs to define properly a phase space. A simple example illustrating how an unshifted function can give rise to a linearly independent one by a simple shift is provided by the polynomials $X_0(t) = t^2$, $X_0(t+\tau) = t^2 + 2t\tau + \tau^2$. These two functions are indeed independent, as

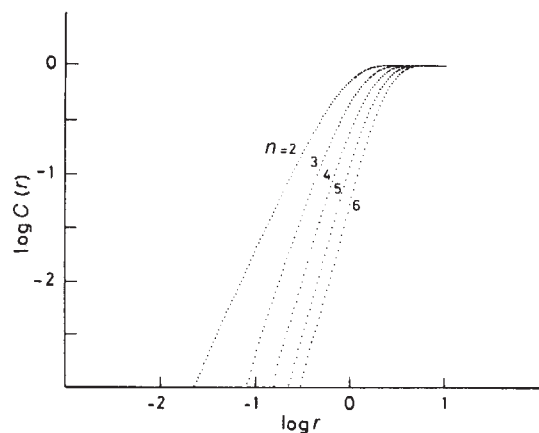


Fig. 4 Distance dependence of the correlation function on the climatic attractor. Note the existence of an extensive range of linearity, from which the dimensionality of the attractor can be inferred. Parameter values as in Fig. 3, except that $\tau = 8$ kyr.

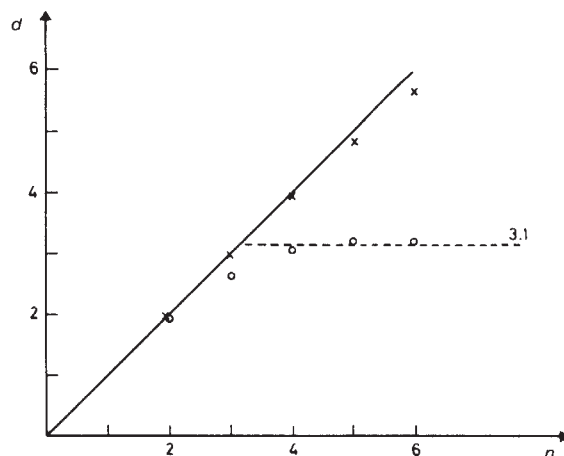


Fig. 5 Dependence of dimensionality, d , on the number of phase space variables, n for the climatic attractor ($\circ$) and for a white noise signal ($\times$) for the same parameters as in Fig. 4. Notice the saturation to a plateau value of ~ 3.1 in the first case, and the $d \sim n$ relationship in the second case.

the second polynomial contains terms of first and zero degree in t , which are absent in the first one.

Before describing the procedure that will enable us to obtain quantitative information on the nature of the climatic attractor and, in particular, on its dimensionality we present, in Fig. 3, a characteristic view of this object in a three-dimensional phase space. This view clearly exhibits the complexity of the underlying motion. However, the data are too coarse to draw any conclusion from this figure alone. It will be our task to characterize the complexity of the dynamics more sharply and, in particular, to assess its similarities and differences with random noise. Note that similar phase portraits have been constructed from experimental data pertaining to chemical kinetics¹⁰.

Analytical procedure

We now describe briefly a recently proposed procedure^{11,12} which allows the identification of some salient features of the phase space portrait of a dynamical system. Consider a set of N points on an attractor embedded in a phase space of n dimensions, obtained from a time series:

$$\begin{aligned} &X_0(t_1), \dots, X_0(t_N) \\ &X_0(t_1 + \tau), \dots, X_0(t_N + \tau) \\ &\vdots \\ &X_0(t_1 + (n-1)\tau), \dots, X_0(t_N + (n-1)\tau) \end{aligned} \quad (2)$$

For convenience we introduce a vector notation: X_i which stands for a point of phase space whose coordinates are $\{X_0(t_i), \dots, X_0(t_i + (n-1)\tau)\}$.

A reference point X_i from these data is now chosen and all its distances $|X_i - X_j|$ from the $N-1$ remaining points are computed. This allows us to count the data points that are within a prescribed distance, r from point X_i . Repeating the process for all values of i , one arrives at the quantity

$$C(r) = \frac{1}{N^2} \sum_{i,j=1}^N \theta(r - |X_i - X_j|) \quad (3)$$

where θ is the Heaviside function, $\theta(x) = 0$ if $x < 0$, $\theta(x) = 1$ if $x > 0$. The non-vanishing of this quantity measures the extent to which the presence of a data point X_i affects the position of the other points. $C(r)$ may thus be referred to as the (integral) correlation function of the attractor.

Suppose that we fix a given small parameter ε and we use it to define the site of a lattice which approximates the attractor. If the latter is a line, the number of data points within a distance

r from a prescribed point should be proportional to r/ε . If it is a surface, this number should be proportional to $(r/\varepsilon)^2$ and, more generally, if it is a d -dimensional manifold it should be proportional to $(r/\varepsilon)^d$. We expect, therefore, that for r , relatively small $C(r)$ should vary as

$$C(r) = r^d \quad (4)$$

In other words, the dimensionality d of the attractor is given by the slope of the $\log C(r)$ versus $\log r$ in a certain range of values of r :

$$\log C(r) = d |\log r| \quad (5)$$

This property remains valid for attractors of fractal dimensionality.

The above results suggest the following algorithm (a similar procedure has also been proposed in the context of fluid dynamics^{13,14}):

(1) Starting from a time series, we can construct the correlation function, equation (3), by considering successively higher values of the dimensionality n of phase space.

(2) Deduce the slope d near the origin according to equation (5) and see how the result changes as n is increased.

(3) If d reaches a saturation limit beyond some relatively small n , the system represented by the time series should possess an attractor. The saturation value d_s will be regarded as the dimensionality of the attractor. The value of n beyond which saturation is observed will provide the minimum number of variables necessary to model the behaviour represented by the attractor.

Application to climatic data

This procedure has been applied to the analysis of the data pertaining to core V28-238. Figure 4 gives the dependence of $\log C(r)$ versus $\log r$ for $n = 2$ to $n = 6$. We see that there is indeed an extended region over which this dependence is linear, in accordance with equation (5). Figure 5 (points in circles) shows that the slope reaches a saturation value at $n = 4$, which is about $d_s = 3.1$. The same plot also shows the way in which d varies with n if the signal considered is a Gaussian white noise: there is no tendency to saturate. In fact, in this case d turns out to be equal to n . It should be emphasized that the above results are independent of the choice of the time lag τ , provided that the latter is of the order of magnitude of the times scales pertaining to the long-term climatic evolution, and the linear independence of the variables is secured.

The existence of a climatic attractor of low dimensionality shows that the main feature of long-term climatic evolution may be viewed as the manifestation of a deterministic dynamics,

involving a limited number of key variables. The fact that the attractor has a fractal dimensionality provides a natural explanation of the intrinsic variability of the climatic system, despite its deterministic character¹⁵. Moreover, it suggests that despite the pronounced peaks of spectra in the frequencies of the orbital forcings, the actual behaviour is highly non periodic¹⁶. A new interpretation of variance spectra of the ice volume record is, therefore, necessary.

Conclusion

We have identified a number of intrinsic properties of the climate assuming it to be dynamical system, using only the time series obtained from the data.

Our results do not anticipate the validity of any particular model of climatic evolution. Rather, they set a number of con-

straints that should be satisfied by a model. In particular, they suggest that models involving four variables could already provide a description of the salient features of the system.

Our approach could be applied to many other problems in which naturally occurring complex systems are probed through time series. Within the context of atmospheric physics and climatology, one important example is the blocking transition. Another promising field of application are biological rhythms like the heart beat or the electroencephalogram¹⁷. In the long run, it may be hoped that fractal dimensionality could become for such systems a useful characterization of their specificity as well as a measure of their complexity.

We thank P. Pestiaux for providing the time series of the V28-238 core and P. Bergé for an enlightening discussion. The work of C.N. is supported, in part, by the EEC under contract no. STI-004-J-C(CD).

Received 2 May; accepted 27 July 1984.

1. Berger, A. L. (ed.) *Climatic Variations and Variability: Facts and Theories* (Reidel, Dordrecht, 1981).
2. Mandelbrot, B. *Fractals: Form, Chance and Dimension* (Freeman, San Francisco, 1977).
3. Helleman, R. (ed.) *Ann. N.Y. Acad. Sci.* **347**, 1981.
4. Shackleton, N. J. & Opdyke, N. D. *Quat. Res.* **3**, 39–55 (1973).
5. Shackleton, N. J. *et al. Nature* **307**, 620–623 (1984).
6. Imbrie, J. *et al. in Milankovitch and Climate* (eds Berger, A. L. *et al.*) (Reidel, Dordrecht, 1984).
7. Berger, A. L. & Pestiaux, P. *Tech. Rep. No. 28*, (Institut of Astronomy and Geophysics, Catholic University of Louvain, 1982).

8. Packard, N. H., Crutchfield, J. P., Farmer, J. D. & Shaw, R. S. *Phys. Rev. Lett.* **45**, 712–716 (1980).
9. Ruelle, D. in *Nonlinear Phenomena in Chemical Dynamics*, (eds Pacault, A. & Vidal, C.) (Springer, Berlin, 1981).
10. Roux, J. C., Simoyi, R. H. & Swinney, H. L. *Physica* **8D**, 257–266 (1983).
11. Grassberger, P. & Procaccia, I. *Physica* **9D**, 189–208, 1983.
12. Grassberger, P. & Procaccia, I. *Phys. Rev. Lett.* **50**, 346–349 (1983).
13. Brandstätter, A. *et al. Phys. Rev. Lett.* **51**, 1442–1445 (1983).
14. Malmanson, B., Atten, P., Berge, P. & Dubois, M. *J. phys. Lett.* **44**, 897–902 (1983).
15. Lorenz, E. N. *Tellus* **36A**, 98–110 (1984).
16. Le Treut, H. & Ghil, M. *J. geophys. Res.* **88**, 5167–5190, 1983.
17. Babloyantz, A., Nicolis, G. & Nicolis, J. S. (in preparation).

Structure of the nucleosome core particle at 7 Å resolution

T. J. Richmond, J. T. Finch, B. Rushton, D. Rhodes & A. Klug

MRC Laboratory of Molecular Biology, University Postgraduate Medical School, Hills Road, Cambridge CB2 2QH, UK

The crystal structure of the nucleosome core particle has been solved to 7 Å resolution. The right-handed B-DNA superhelix on the outside contains several sharp bends and makes numerous interactions with the histone octamer within. The central turn of superhelix and H3·H4 tetramer have dyad symmetry, but the H2A·H2B dimers show departures due to interparticle associations.

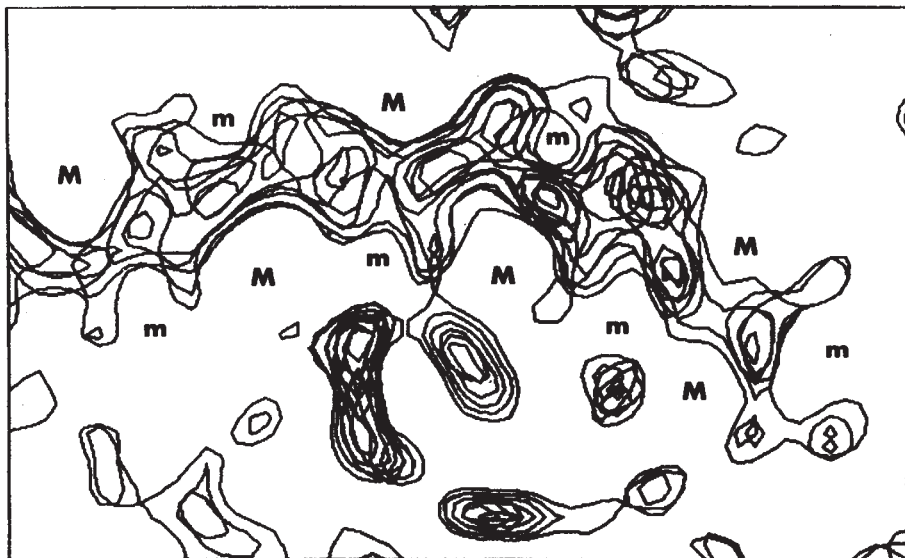
THE nucleosome is the primary repeating unit of DNA organization in chromatin^{1–3}. Extensive digestion of chromatin with micrococcal nuclease releases the nucleosome core, a small, well-defined particle which has been crystallized. The particle mass (~206 kilodaltons) is equally distributed between 146(±2) base pairs (bp) of DNA and an octamer formed by two each of the histone proteins, H2A, H2B, H3 and H4 (the outer histone H1 and the linker DNA having been removed during nuclease digestion⁴). The crystals first obtained were of limited order and gave a picture of the particle to 20 Å resolution⁵. We later prepared better crystals which diffracted to about 5 Å, having a single particle in the asymmetric unit⁶. We describe here the three-dimensional structure of the nucleosome core particle at 7 Å resolution determined by X-ray diffraction and multiple isomorphous replacement. This advance was achieved by improved diffractometry, control of the crystal unit cell parameters by adjustment of hydration, and derivatives prepared from multi-heavy atom cluster compounds.

The nucleosome core crystals have previously been investigated by several structural methods. Electron microscopy and X-ray diffraction techniques were initially used to obtain electron density maps in the principal projections, suggesting that the particle had the shape of a disk 57 Å thick and 110 Å in diameter. The DNA appeared to be wound around the histone

core in 1.8 turns of a flat superhelix with a pitch of 28 Å⁵. Using neutron diffraction in conjunction with contrast matching^{7,8}, the DNA and protein components were seen as separate densities in projections at 25 Å resolution. The three-dimensional shape of the octamer alone was determined to a resolution of 20 Å by image reconstruction from electron micrographs of negatively stained, ordered aggregates⁹. The assignment of histone locations on the resulting low-resolution map relied on DNA-histone^{10,11} and histone-histone (for references, see ref. 12) proximities obtained from chemical cross-linking experiments, and the neutron experiments fixed the orientation of the DNA superhelix on the octamer. The suggested division of the octamer into two H2A·H2B dimers residing on opposite faces of a H3·H4 tetramer agrees well with the pattern of dissociation found in earlier physicochemical studies^{13,14}. The tetramer (proteolysed) and octamer have also been crystallized in the absence of DNA, but structures have not yet been reported^{15,16}.

The electron density map at the higher resolution of 7 Å described here reveals many new structural features. The DNA is not bent uniformly into the superhelix, but exhibits several regions of tight bending or possible kinking, adjacent to points of substantial contact with histones H3 and H4. The histone-DNA interactions occur on the inside of the superhelix; the protein density does not seem to embrace the outside of the

Fig. 1 Electron density of an arc of the DNA superhelix, showing two turns of the double helix. Alternating major (M) and minor (m) grooves are indicated. Three sections (-10 to -12) of the map, 1.73 Å apart, were stacked to show a slice through the double helix. The density seen below the arc of DNA is part of H2B¹.



DNA nor protrude between the turns of the superhelix. Protein domains are now visible, consistent with the earlier histone assignments, and the location of histone H3 has been confirmed by a heavy atom label. Non-crystallographic dyad symmetry relates the halves of the H3·H4 tetramer and the central turn of DNA superhelix; the dyad symmetry, however, does not extend to the same degree to the two H2·H2B dimers, due to interactions between adjacent particles in the crystal.

X-ray analysis and map interpretation

Nucleosome cores from beef kidney were prepared and crystallized as described previously⁶. The space group is $P2_12_12_1$, with one particle of 206,000 daltons per asymmetric unit. The unit cell dimensions of the crystals are dependent on the mean length of the core DNA in a particular preparation^{6,17}. The *b*-axis dimension shows the greatest variability, about 4 Å per base-pair increment. The cell dimensions also depend on the degree of hydration, and we found that the crystalline order was increased by shrinkage to an essentially stable optimum state. Thus in 20% 1,6-hexanediol, the *b*-axis length decreases from 200 Å to 184 Å, and the order increases by about 1 Å in resolution in all directions, although the fall-off in intensity remains anisotropic. As a test for particle integrity, DNase I digestion was carried out in the presence of hexanediol, and the positions and frequencies of cutting were found to be unaffected except for minor deviations near the ends of the DNA. Naked DNA would be unwound by about 3% in screw angle at the alcohol concentrations used here¹⁸. No change in the high angle X-ray pattern⁶ characteristic of B-form DNA was detected.

A variation of $\pm 1.5\%$ in the *b*-cell dimensions between different crystals was compensated by adjusting the alcohol concentrations over the range 21–27% to systematically obtain cell dimensions of $a = 107.5$, $b = 184$, $c = 110.5$ Å. The crystal was mounted in a solution containing low gelling temperature agarose which set on cooling to maintain constant hydration during data collection at 4°C and to keep it fixed in position. The glass capillaries were gently tapered at one end to provide a good fit to the rod-shaped crystals, thereby minimizing X-ray absorption; this was less overall in this method than in conventional mounting techniques, and resulted in a deviation of typically $\pm 5\%$ about the capillary axis.

X-ray diffraction data were collected with a modified four-circle, Hilger–Watts diffractometer mounted on an Elliot rotating anode GX13 X-ray source, using well collimated beams to ensure accurate data at low resolution. For each dataset, the 3,800 reflections from 35 Å to 7 Å and their Friedel mates were measured twice using a partial peak scan method followed by

profile fitting analysis to yield the integrated intensities (T.J.R., in preparation). In total, 64 crystals were used for the three datasets. The merging *R*-factors for the amplitudes were 0.05–0.06 over all crystals and symmetry equivalents for a dataset. A crystal was abandoned after one of several reference reflections had decayed by 20% of its initial intensity.

Two heavy atom derivatives were used to determine the phases by the method of multiple isomorphous replacement. Cluster compounds containing several heavy atoms were used in order to overcome the difficulties imposed by the large asymmetric unit and fall-off in the diffraction intensity. The first was tetrakis(acetoxymethyl)mercury (TAMM)¹⁹ used at 2 mM and solubilized in 20 mM glycylglycine. TAMM bound to essentially a single site as determined from difference Patterson projections. The second heavy atom cluster, di- μ -iodo, bis(diaminoethyl-*N,N'*)diplatinum(II) (PIP) was used at 0.1 mM and found to bind at three major and several minor sites which were located by difference Fourier projections using the phases generated from the TAMM derivative. The *R*-factors for the scaling of the TAMM and PIP datasets to the native dataset were 0.12 and 0.17, respectively. The heavy atom parameters were refined against the centric data using the F_{hlc} procedure²⁰, and resulted in phasing power factors, f_{H}/E , for the two derivatives of 1.55 and 1.72, respectively (f_{H} is the r.m.s. calculated heavy atom contribution to the structure factor, and E is the r.m.s. closure error). The figures of merit over all the data were 0.543 for the selected enantiomorph and 0.529 for the opposite hand.

The major and minor grooves of the DNA (or conversely the phosphodiester backbones) are clearly visible as alternating cusps in the electron density map over all the expected length of the DNA, thus defining the DNA superhelix (Fig. 1). Several rods of strong density (presumably α -helices) are seen on the inside of the DNA superhelix, each belonging to discrete regions of electron density, which appear to define the histone domains in a manner consistent with the assignments deduced from the low-resolution map⁹ and chemical cross-linking data^{10–12}. Thus the order of the main histone contacts from end to end of the DNA superhelix is: H2A, H2B, H4, H3, H3, H4, H2B, H2A. The density at the interfaces between histones cannot, however, be assigned unambiguously at this resolution, and we have ignored any dovetailing that might exist, particularly between H3 and H4. The particle dyad axis (see below) was determined by comparing prominent features in the histone tetramer and the grooves of the central turn of the DNA superhelix.

As a check on the assignments, the volumes of the DNA and of the individual histone domains were estimated by summing the relevant contour areas over all sections. The fractional volume of the crystal occupied by the core particle is 0.35 as

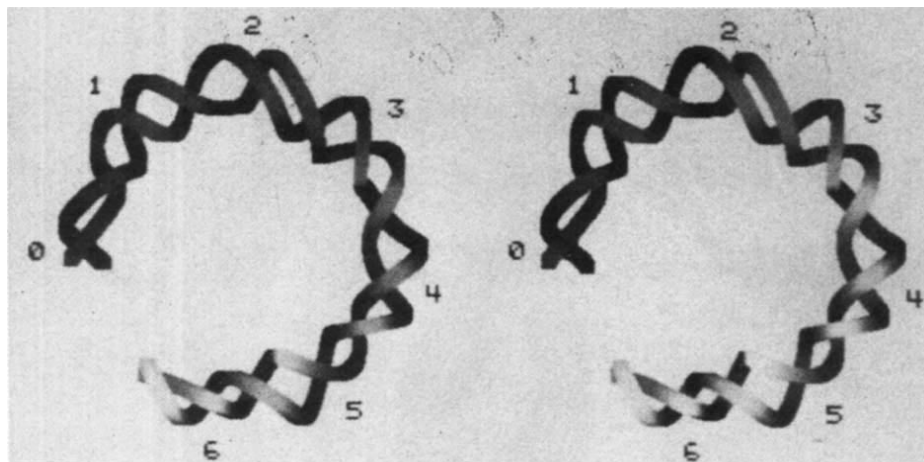


Fig. 2 The folded DNA double helix represented by the course of the phosphodiester backbones. Only one turn of the DNA superhelix as viewed down the superhelix axis is shown for clarity (as a stereopair). The points of intersection of the phosphodiester chains with each map section were joined by cubic spline fitting⁵⁶ to produce the two smooth curves shown. The superhelix is oriented such that the molecular dyad axis is horizontal and passes through the minor groove facing outward at position 0. The other positions where the minor groove faces outward are numbered accordingly. The same orientation of the dyad axis is used in Figs 2-6.

compared with a calculated value of 0.40. This agreement is as good as can be expected, as the contour level (55% of the difference between maximum and minimum peak heights) which should be chosen for the boundaries will depend on unknown factors such as small variations in particle position and orientation and local disorder in the histone N-terminal segments²¹. The counterions associated with the DNA and protein would also affect the partial specific volumes.

The core particles are pseudo-hexagonally packed in layers as described previously^{6,17}. Most of the contacts between particles are made by DNA-DNA interactions, with one important exception. The H2A·H2B dimers on each flat face of the particle associate with the DNA of an adjacent particle in a head-to-tail fashion along the *b*-direction in the layers. Presumably nuclease trimming of the nucleosome has freed a DNA binding site on each dimer, making it available for interparticle associations. There is also a protein-protein interaction (H3²-H2B¹) between layers of the crystals.

The DNA double helix and superhelix

The DNA double helix is in a right-handed B-form (see also ref. 6) and folded into a left-handed superhelix (Fig. 2). About 12 turns of the DNA double helix are clearly visible; the last half turn on each end is indistinct owing to the variable length of the DNA and crystal packing effects. As before⁹, we mark the distance along the DNA by numbers from -7 to +7 (from end to end), taking as origin the point where the dyad axis passes through the double helix at the middle of the nucleosomal DNA. The minor groove faces directly outward from the particle at this point, position 0 (Fig. 3a), which corresponds to site 7 of the DNase I cutting pattern^{22,23}. The positions of the integral numbers correspond approximately to the sites of attack (the local maxima in the frequency of cutting) by the enzyme DNase I²⁴. These sites consist of pairs of cuts staggered by 1-3 bp on the two DNA strands²⁵.

The DNA has on average about 7.6 turns of double helix in a complete superhelical turn. As a result of this periodicity, the adjacent turns of the superhelix are arranged with a minor groove approximately opposite a major groove. The pitch of the superhelix is ~30 Å for the central turn measured between positions -3.8 and +3.8, but reaches a minimum value of 25 Å between positions -2.5 and +5.0. The pitch of the double helix measured along an arc of the superhelix is about 35 Å. At the resolution described here, the actual number of nucleotides per double helical turn cannot be determined. The precise nature of the distortions in the DNA should be apparent when the data to 5 Å resolution are included, as the phosphates should then be visible.

The DNA double helix does not take up a perfectly regular superhelical path as it winds around the histone octamer, but is bent fairly sharply or possibly kinked at several locations

(Fig. 2). The DNA bend in the region of position +1 is itself not in direct contact with the histones, but bridges between an extensive histone interaction at position +0.5 on one side and lesser contacts on the other. A similar bend occurs at position +4, apparently due to the DNA-histone contact at position +3.5. The symmetrically related positions at -1 and -4 show similar DNA bending. These regions of high curvature are apparently unlike previously proposed kinks (for example, ref. 26), as the departure from good stacking is apparently spread over several base pairs. Despite the evident ovaloid shape of the superhelix as viewed down its axis (Fig. 2), its diameter (measured between axes of the double helix across the particle) is fairly constant at 86 ± 4 Å.

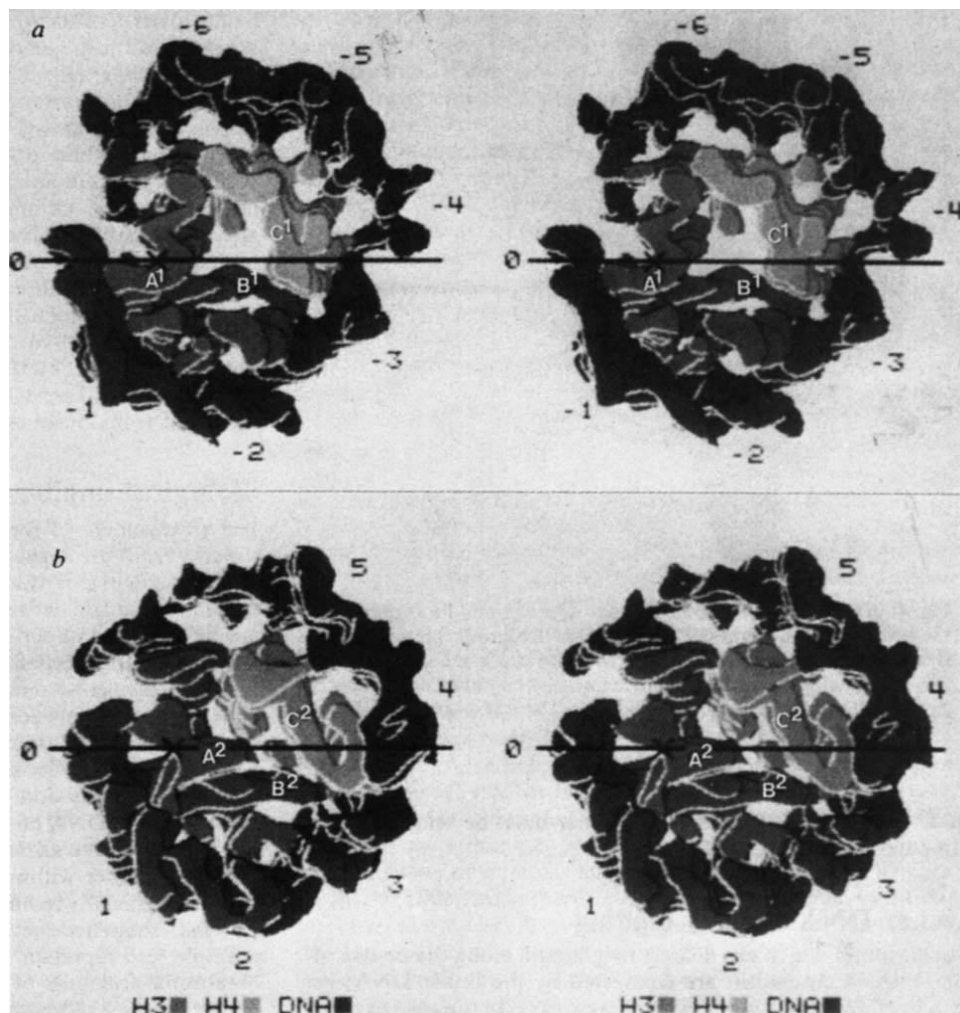
Because the widths of the major and minor groove seemed to depend on the degree of bending (Fig. 2), we measured them at both the inside and outside surfaces of the double helix at several sites which are exceptionally clear in the map. At position +1 for example, the minor groove width increases from 7 Å on the inside of the particle to 13 Å on the outside. Similarly, the major groove widths are 11 Å and 20 Å. The major groove appears to be more compressed than the minor groove on the inside of the most highly curved regions.

The positions of tightest DNA bending (± 1 , ± 4) correlate well with the sites of greatest 'protection' from DNase I digestion²⁴ and enhanced reactivity with dimethyl sulphate²⁷. Because no obvious protein density is seen extending around the outside of the DNA, it appears that the reduced nuclease cleavage rates do not arise through simple steric hindrance, but rather from distortions in the DNA conformation, or from a reduction of its flexibility due to tight binding by the histones. It has been shown for a double helical oligonucleotide of defined sequence that the cleavage rate for DNase I depends on the local variation in the screw of the double helix^{28,29}. Conformational differences along the nucleosome superhelix might indeed lead to relative enhancements of nuclease cleavage rate and also of chemical reactivity. Dimethyl sulphate labelling of purines shows altered reactivity at only two bases as compared with free DNA²⁷, and these are located in the inward-facing major groove free of protein at positions ± 1 , where the double helix is sharply bent, and in the minor groove in contact with histone at ± 0.5 .

The histone octamer and DNA interactions

As mentioned above, features in the protein part of the map are clustered and correspond to the eight individual histone molecules. The assignment of histone H3 is confirmed by the location of the major binding site for the TAMM heavy atom cluster. This label falls precisely on the particle dyad axis (Fig. 3a), suggesting that this tetravalent compound bridges between two invariant sulphhydryl groups, one from each H3. These are known to be close enough to form a disulphide bond³⁰. Even though there are two -SH groups in each bovine H3, we can

Fig. 3 The two halves of the H3·H4 tetramer and the central turn of the DNA superhelix, both viewed from inside outwards. The map is presented in the form of a stereo-pair of a computer-generated model and has here been sliced into two 15 Å thick slabs parallel to the flat faces of the core particle: *a*, sections -12 to -3; *b*, sections -1 to 6. The alternating groove structure of the DNA is seen particularly clearly from positions 0 to +4 in *b*. The half-tetramer in *b* has been rotated 180° about the dyad axis to give the same view as in *a*; thus the arc 1 to 4 fits against the arc -7 to -3 in the other half. The position of the molecular dyad axis was determined from the rods of histone density labelled B¹ and C¹, and B² and C². Some of the contours in the histone core have been removed to allow these rods to be seen. The TAMM heavy atom cluster site on the H3 dimer is indicated by a bold X.



be certain that cysteine 110 is the labelled group, because the level of mercurial binding and the intensity distribution in the X-ray patterns from TAMM-reacted crystals of bovine core particles are the same as those from chicken erythrocytes (data not shown), which contain only the two cysteine 110 residues. As can be seen in Figs 3 and 4, the heavy atom site lies between two rods of electron density ~ 20 Å long (A¹ and A²) which about the minor grooves of the DNA at sites -0.5 and +0.5. The H3 densities fold back from the DNA at these positions and extend across the particle for ~ 35 Å in long rods of density (B¹ and B²).

Two other rod-like features ~ 35 Å in length (C¹ and C²) are visible in the same map sections, but appear to be disconnected from the features defining the H3 region. Therefore, they are likely to be part of the H4 molecules. Shorter pieces of density fold back from the ends of these rods, and make substantial contact with the DNA minor grooves at positions -3.5 and +3.5. These grooves lie on either side of the particle dyad axis at the point where it passes between the two turns of the superhelix. Histones H4 also appear to make minor contacts with the outer superhelical turns of DNA from positions ± 4 to ± 5 .

The two pairs of rod-like density, B¹C¹ and B²C², prominent in the tetramer region, were aligned to determine the orientation of the non-crystallographic dyad symmetry axis (Fig. 3). Other features in the H3·H4 histone tetramer and the central turn of the DNA superhelix are well related by this dyad axis.

DNA-protein contacts are made on nearly every turn of the DNA double helix and are confined to the inner surface of the superhelix. If each such histone contact contained a positive charge, the degree of phosphate charge neutralization thereby afforded would be consistent with the value of 20% derived

from the analysis of melting of core particle DNA³¹. If any parts of the histones larger than, say, an extended chain are bound to the outer surface of the DNA, they are disordered in the crystals because there are no places in the 7 Å map where the DNA grooves on the outside of the particle are obscured by protein density. The spaces between the turns of the superhelix are also free from obvious protein density.

In contrast to the tetramer, the pair of H2A·H2B dimers clearly depart from dyad symmetry in the crystals. The whole core particle forms the asymmetric unit, so that the two copies of each histone are not constrained to identical environments. Being on the outer surface of the particle, the H2A·H2B dimers are more susceptible to disturbance by crystal packing forces than is the internally located H3·H4 tetramer.

The H2A¹·H2B¹ dimer is a striking feature of the electron density map. The H2B¹ domain contains an ~ 40 Å long rod of density, D¹ (Fig. 5). Contacts with minor grooves are made at positions -3.5, -4.5 and -6.5 whereas the protein density runs roughly parallel with the major groove at -5. The H2A¹ domain extends over the DNA at position 0 to the outer edge of the core particle, and is bound to the major groove at site -3.5 of the DNA on the neighbouring particle (Fig. 5).

The second copy of the dimer is less well ordered and has apparently turned (anticlockwise from the position in Fig. 6) about a point in H2B² such that the body of H2A² is shifted by ~ 10 Å from the position which is symmetrically equivalent to H2A¹. This departure from symmetry may represent an intermediate position of this dimer during a quaternary structural change relative to the H3·H4 tetramer. Like H2A¹, H2A² is associated with the DNA at position -7 on an adjacent particle in the crystal. In order to make this DNA interaction, the

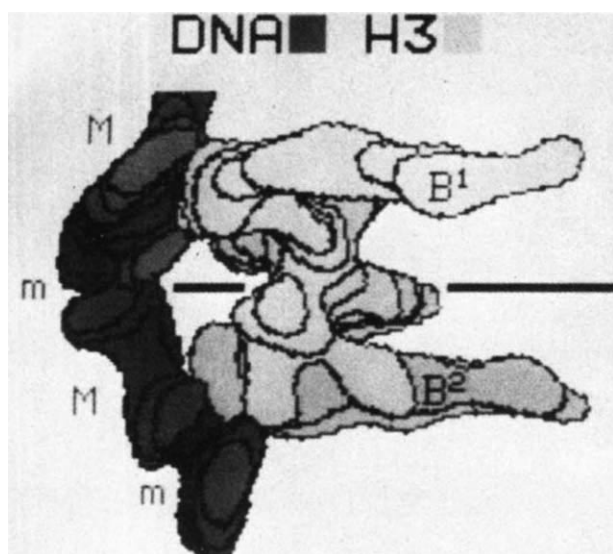


Fig. 4 The H3 dimer and its contact with the middle region of the superhelix. The terminal regions of the long rods, presumably α -helices, make extensive contacts with the minor grooves of the DNA one-half turn of the double helix away on either side of the dyad axis, represented by the solid line. The minor groove faces away from H3 on the dyad.

in the particle and do not protrude over position 0, but this would entail a major distortion on the diametrically opposite side of the superhelix at position ± 4 , which we consider unlikely.

This proposed arrangement of the DNA one double helical turn beyond the observed ends of the core particle DNA suggests how H1 may be bound on the nucleosome. Electron microscopy studies of H1-containing and H1-depleted chromatin³³ and the isolation of the chromatosome (which contains H1 and an additional 20 bp of DNA³⁴) indicate that H1 is associated with both ends of the superhelix as they make the transition into the linker region. The three DNA segments -7 to -8 , $+7$ to $+8$, and -1 to $+1$ would form, when viewed edge-on down the dyad axis, a three-sided cavity, which would provide a reasonable binding site for the H1 globular domain³⁵. The core histones that would be closest to H1 are the H2A molecules, as was indicated from chemical cross-linking studies^{36,37}.

Biological implications

The arrangement of the histones and their degrees of contact with the DNA are suggestive of the roles of the different histones in the assembly of the nucleosome^{9,17}. The most substantial contacts are made between the H3·H4 tetramer and the central turn of the DNA superhelix, and, as was foreseen⁹, account for the fact that the tetramer alone can protect about 80 bp of DNA against cleavage by micrococcal nuclease^{38,39}. Weaker interactions also occur between the tetramer and the outer turns of the superhelix and explain how the tetramer alone can also organize ~ 140 bp into a loose nucleosome-like particle^{40,41}. The H2A·H2B dimers add to each exposed face of the tetramer, binding to the DNA of the last half turns of the superhelix and increasing the overall stability of the nucleosome.

The H3 dimer within the tetramer binds to DNA in a way which superficially resembles the DNA binding model suggested for the sequence-specific proteins from prokaryotes (for example, cro repressor⁴², catabolic activator protein⁴³ and the N-terminal fragment of λ cI repressor⁴⁴). These proteins form dimers whose dyad symmetry axis could align with a local dyad axis in the DNA structure as we have found for the H3 dimer, but the details may differ. For the prokaryotic proteins, a pair of α -helices may contact the edges of the bases in the major grooves of the DNA one half turn to either side of the dyad, whereas, in the case of the H3 dimer, the ends of presumptive α -helices lie against the minor grooves of the DNA one-half turn of the double helix away from the central dyad axis (Fig. 4). Another variation is seen in the complex of the EcoRI restriction enzyme and its specific DNA recognition sequence. The protein and DNA are oriented as for the H3 dimer-DNA interaction, but dimerically related α -helices contact the bases in an expanded major groove facing inward on the dyad⁴⁵.

H2A²·H2B² contacts with the tetramer must be rearranged as compared with H2A¹·H2B¹.

Linker DNA and H1 binding

Nucleosomes which are nearest neighbours along the course of the DNA in chromatin are connected by the linker DNA; its path in chromatin is unknown and may vary. In the core particle structure determined here, the spaces to either side of the DNA at position 0 are blocked by H2A molecules (Fig. 6). If this is also true in solution, the linkers, which would in a nucleosome extend beyond positions -7 and $+7$, could not simply follow the smooth superhelical course set by the central turn of DNA, as has been suggested^{5,32}. The H2A molecules extend to the outer edge of the particle, and are likely to cause the linkers to make angles of roughly 50° to 60° with the molecular dyad axis. Chromatin containing very short linker lengths (that is, 0–10 bp) would then require the DNA to bend sharply or kink, perhaps as seen on the core particle itself. It could be argued that in solution the H2A·H2B dimers take up a more central position

Fig. 5 The H2A¹·H2B¹ dimer and outer half turn of the DNA superhelix, viewed from the inside. The stretch of DNA from its end position at -7 to position -3.5 is shown: the groove structure in the region -4 to -5 is very clear. The H2B¹ domain includes a number of distinct features such as the partly buried rod of density (D¹) crossing the length of the domain, and a piece of density running roughly in the major groove at position -5 . The point on the H2A² domain which contacts the DNA of the adjacent particle (in the horizontal direction) at position -3.5 is indicated by the X. This contact is repeated along the crystallographic two-fold screw axis in the b -direction.

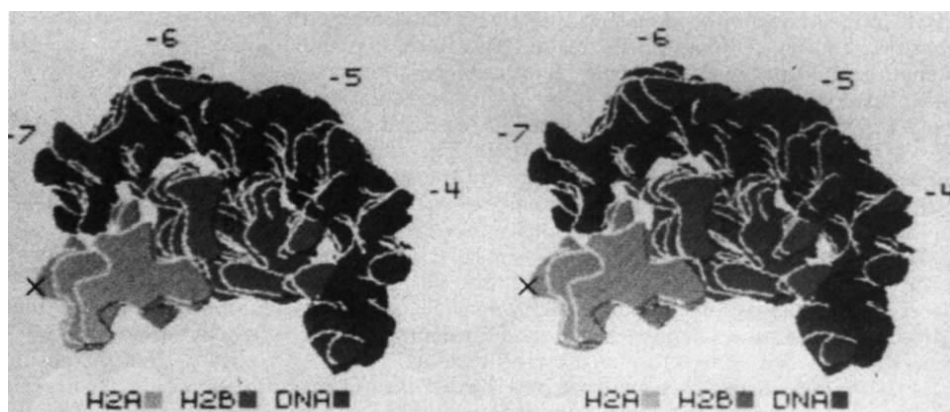
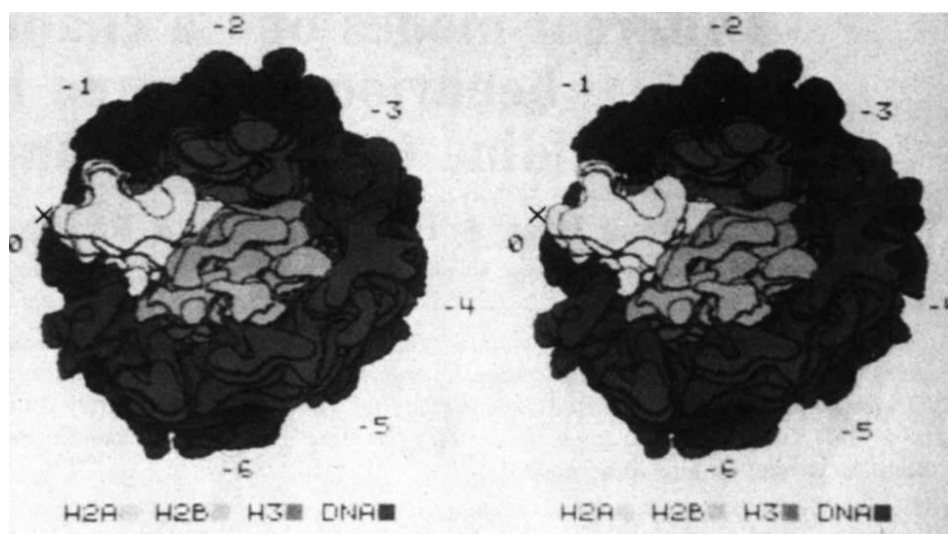


Fig. 6 The location of the H2A¹·H2B¹ dimer in the nucleosome core particle. The entire core particle is shown (sections -21 to +13) to illustrate that H2A¹ is positioned over the central turn of the DNA superhelix near position 0. The point marked X is the same as in Fig. 5.



The sharp bends in the DNA path on the nucleosome suggest a structural explanation for the observation that certain DNA sequences are specifically positioned on the histone octamer^{46,47}, a phenomenon that has been called nucleosome phasing⁴⁸ (but see ref. 49). It may be that on account of their relative flexibility, certain short nucleotide sequences are preferentially accommodated at the sharp bends, whereas others may be preferentially excluded^{50,51} because of their stiffness. The interplay of these preferences in different regions of a DNA sequence of roughly nucleosome length could result in fixing it at a favoured position. This mode of positioning would not depend directly on specific interactions between the histones and certain base pairs, although it does not, of course, exclude them.

The compact structure that the genomic DNA attains in the nucleosome is essential for chromosome segregation during mitosis, yet the nucleosome cannot be a static structure throughout the cellular processes of transcription and replication. Several recent studies indicate that the histone octamer of transcriptionally active chromatin dissociates from the DNA⁵²

or takes on an altered structure^{53,54}. In the crystal structure, one of the H2A·H2B dimers appears to have moved due to an interaction of H2A with the DNA of another particle. This suggests how the nucleosome might disassemble in response to an alteration in the path of the linker DNA joining nucleosomes in chromatin, perhaps in response to the passage of DNA or RNA polymerase. It seems possible that the dimer would dissociate from the tetramer, while both oligomers maintain separate contact with the DNA. The movement of the H2A·H2B dimer with respect to the histone tetramer is of particular interest in view of the report of a complex formed between RNA polymerase II and nucleosome core particles enriched in transcribed DNA sequences, because this complex lacks one of the dimers⁵⁵.

We thank S. J. Lippard and T. O'Halloran for their gifts of the TAMM and PIP heavy atom cluster compounds, the preparation and use of which will be described elsewhere in a joint communication (S.J.L., T.O'H., T.J.R. and A.K., in preparation). We dedicate this paper to Dr Max Perutz on the occasion of his 70th birthday.

Received 30 May; accepted 27 July 1984.

- Kornberg, R. D. *A. Rev. Biochem.* **46**, 931-954 (1977).
- McGhee, J. D. & Felsenfeld, G. *A. Rev. Biochem.* **49**, 1115-1156 (1980).
- Klug, A. *Symposium on Nucleic Acid Res.* (eds Mizobuchi, K., Watanabe, I. & Watson, J. D.) 91-112 (Academic, New York, 1983).
- Noll, M. & Kornberg, R. D. *J. molec. Biol.* **109**, 393-404 (1977).
- Finch, J. T. *et al. Nature* **269**, 29-36 (1977).
- Finch, J. T. *et al. J. molec. Biol.* **145**, 757-769 (1981).
- Finch, J. T., Lewit-Bentley, A., Bentley, G. A., Roth, M. & Timmins, P. A. *Phil. Trans. R. Soc. B* **290**, 635-638 (1980).
- Bentley, G. A., Finch, J. T. & Lewit-Bentley, A. *J. molec. Biol.* **145**, 771-784 (1981).
- Klug, A., Rhodes, D., Smith, J., Finch, J. T. & Thomas, J. O. *Nature* **287**, 509-516 (1980).
- Mirzabekov, A. D., Shick, V. V., Belyavsky, A. V. & Bavykin, S. G. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4184-4188 (1978).
- Shick, V. V., Belyavsky, A. G., Bavykin, S. G. & Mirzabekov, A. D. *J. molec. Biol.* **139**, 491-517 (1980).
- Thomas, J. O. & Kornberg, R. D. *FEBS Lett.* **58**, 353-358 (1975).
- Kornberg, R. D. & Thomas, J. O. *Science* **184**, 865-868 (1974).
- Eickbush, T. H. & Moudrianakis, E. N. *Biochemistry* **17**, 4955-4964.
- Lattman, E., Burlingame, R., Hatch, C. & Moudrianakis, E. N. *Science* **216**, 1016-1018 (1982).
- Burlingame, R. & Moudrianakis, E. N. *Science* **216**, 1016-1018 (1984).
- Lipka, J. J., Lippard, S. J. & Wall, J. S. *Science* **206**, 1419-1421 (1979).
- Dodson, E. J. in *Crystallographic Computing Techniques* (ed. Ahmed, F. R.) 259-268 (Munksgaard, Copenhagen, 1976).
- Cary, P. D., Moss, T. & Bradbury, E. M. *Eur. J. Biochem.* **89**, 475-482 (1979).
- Lutter, L. C. *J. molec. Biol.* **117**, 53-69 (1977).
- Sollner Webb, B. & Felsenfeld, G. *Cell* **10**, 537-547 (1977).
- Lutter, L. C. *J. molec. Biol.* **124**, 391-420 (1978).
- Lutter, L. C. *Nucleic Acids Res.* **6**, 41-56 (1979).
- Crick, F. C. & Klug, A. *Nature* **255**, 530-533 (1975).

- McGhee, J. D. & Felsenfeld, G. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2133-2137 (1979).
- Lomonosoff, G. P., Butler, P. J. G. & Klug, A. *J. molec. Biol.* **149**, 745-760 (1981).
- Dickerson, R. E. & Drew, H. R. *J. molec. Biol.* **125**, 535-543 (1981).
- Camerini-Otero, R. D. & Felsenfeld, G. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5519-5523 (1977).
- McGhee, J. D. & Felsenfeld, G. *Nucleic Acids Res.* **8**, 2751-2769 (1980).
- McGhee, J. D., Nichol, J. M., Felsenfeld, G. & Rau, D. C. *Cell* **33**, 831-841 (1982).
- Thoma, F., Koller, Th. & Klug, A. *J. Cell Biol.* **83**, 403-427 (1979).
- Simpson, R. T. *Biochemistry* **17**, 5524-5531 (1978).
- Allan, J., Hartman, P. G., Crane-Robinson, C. & Aviles, F. X. *Nature* **288**, 675-679 (1980).
- Boulakas, T., Wiseman, J. M. & Garrard, W. T. *Proc. natn. Acad. Sci. U.S.A.* **77**, 127-131 (1980).
- Thomas, J. O. & Khabaza, A. J. A. *Eur. J. Biochem.* **112**, 501-511 (1980).
- Camerini-Otero, R. D., Sollner-Webb, B. & Felsenfeld, G. *Cell* **8**, 333-347 (1976).
- Stockley, P. G. & Thomas, J. O. *FEBS Lett.* **99**, 129-135 (1979).
- Thomas, J. O. & Oudet, P. *Nucleic Acids Res.* **71**, 611-623 (1979).
- Simon, R. H., Camerini-Otero, R. D. & Felsenfeld, G. *Nucleic Acids Res.* **5**, 4805-4818 (1978).
- Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. *Nature* **290**, 754-758 (1981).
- McKay, D. B. & Steitz, T. A. *Nature* **290**, 744-749 (1981).
- Pabo, C. O. & Lewis, M. *Nature* **298**, 443-447 (1982).
- Frederick, C. A. *et al. Nature* **309**, 327-331 (1984).
- Chao, M. W., Gralla, J. & Martinson, H. G. *Biochemistry* **18**, 1068-1074 (1979).
- Simpson, T. R. & Stafford, D. W. *Proc. natn. Acad. Sci. U.S.A.* **80**, 51-55 (1983).
- Zachau, H. G. & Igo-Kemenes, T. *Cell* **24**, 597-598 (1981).
- Kornberg, R. D. *Nature* **292**, 579-580 (1981).
- Kunkel, G. R. & Martinson, H. G. *Nucleic Acids Res.* **9**, 6869-6888 (1981).
- Prunell, A. *EMBO J.* **1**, 173-179 (1982).
- Ness, P. J., Labhart, P., Banz, E., Koller, Th. & Parish, R. W. *J. molec. Biol.* **166**, 361-381 (1983).
- Andersson, K., Mahr, R., Bjorkroth, B. & Daneholt, B. *Chromosoma* **87**, 33-48 (1982).
- Prior, C. P., Cantor, C. R., Johnson, E. M., Littau, V. C. & Allfrey, V. G. *Cell* **34**, 1033-1042 (1983).
- Baer, B. & Rhodes, D. *Nature* **301**, 482-488 (1983).
- Hayes, J. G. & Halliday, J. *J. Instn Maths. Appl.* **14**, 89-103 (1974).

Different modes of Ca channel gating behaviour favoured by dihydropyridine Ca agonists and antagonists

Peter Hess, Jeffry B. Lansman & Richard W. Tsien

Department of Physiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA

Single cardiac transmembranous Ca channels have three modes of gating behaviour in the absence of drugs, expressed as current records with brief openings (mode 1), with no openings because of channel unavailability (mode 0 or null mode) and with long-lasting openings and very brief closings that appear only rarely (mode 2). The dihydropyridine Ca agonist Bay K 8644 enhances Ca channel current by promoting mode 2, while the Ca antagonists nitrendipine and nimodipine inhibit the current by favouring mode 0.

DIHYDROPYRIDINES such as nifedipine and nitrendipine are among the newest and most interesting of the many agents that modify Ca channel activity in excitable cells^{1,2}. As therapeutic agents, they have already gained wide use in the treatment of cardiovascular disorders³⁻⁶; as specific ligands, they open the way towards eventual biochemical purification and reconstitution of Ca channels⁷⁻¹⁰. Questions about the mechanism of their action have been raised by the discovery of 'Ca agonists', such as Bay K 8644 and CGP 28 392 (refs 11, 12), which increase cardiac contractility and divalent cation influx¹¹⁻¹⁶ even though they are dihydropyridines and thus very similar in structure to nifedipine and other 'Ca antagonists'. How can such closely related compounds produce effects that seem so diametrically opposite? Here we report analysis of Ca agonist and Ca antagonist effects on Ca channel activity of single heart cells and single Ca channels. We find that Bay K 8644 increases the chances that the channel will show a mode of gating marked by long opening and brief closing events. The same type of gating can be seen occasionally in the absence of drug. On the other hand, nitrendipine and nimodipine exert their Ca antagonist effect by increasing the proportion of sweeps where the channel fails to open. The results provide an explanation for partial agonist or antagonist actions; in particular, we find that the effects of nitrendipine include a Ca agonist component.

Methods

Single ventricular cells were obtained from guinea pig or frog hearts by enzymatic dispersion^{17,18}, and whole cell membrane current recordings were made under voltage clamp conditions using suction pipette methods of Lee *et al.*¹⁹ (guinea pig) or

Hamill *et al.*²⁰ (frog). Patch clamp recordings of single Ca channel activity were made from cell-attached patches on guinea pig ventricular cells^{21,22} with patch pipettes containing 110 mM BaCl₂, 10 mM glucose and 10 mM HEPES (titrated with Ba(OH)₂ to pH 7.5). The cells were bathed in a solution containing 10 mM KCl, 140 NaCl, 1 mM MgCl₂, 10 mM glucose and 10 mM HEPES (titrated with NaOH to pH 7.5). In some of the experiments, the external solution also contained 0.5 mM EGTA. Current signals were filtered with an eight-pole Bessel filter with a cut-off frequency (3 dB) of 1 kHz, and sampled at 5 kHz. Automated analysis of the digitized signals was carried out with procedures similar to those previously described^{21,22}. The analysis program disregarded open or closed periods that lasted up until the time of repolarization. The figures show results from patches containing only one channel, as judged by the absence of a multiple conductance level. All experiments were carried out at room temperature (22 °C).

Ca channel current in single cells

The Ca agonist compound Bay K 8644 strongly enhanced Ca channel current in both guinea pig and frog ventricular cells. With 10 mM external Ca, and test depolarizations from -60 mV to +10 or +20 mV, 1 μ M Bay K 8644 increased the peak inward current by 3.0 ± 0.3 fold (s.e.m.) in guinea pig myocytes ($n = 6$) and 4.05 ± 0.52 fold in frog ventricular cells ($n = 7$). Bay K 8644 at 10 μ M gave no greater enhancement than at 1 μ M; half-maximal effects were obtained at concentrations of 30-100 nM. At maximally effective doses, Bay K 8644 increased whole cell Ca channel current within seconds, and in all cases the enhancement was sustained.

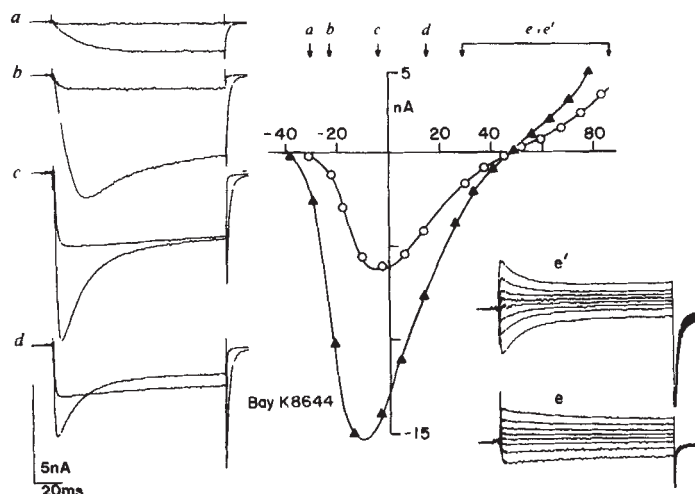


Fig. 1 Voltage-dependent effect of Bay K 8644 on Ca channel currents carried by Ba²⁺. Single guinea pig ventricular cells were studied with a suction pipette for voltage clamp and internal dialysis¹⁹. Each panel shows current signals evoked by step depolarizations from a holding potential of -60 mV, in the absence of drug (smaller peak current, more rapidly decaying tail current), and in the presence of 1 μ M Bay K 8644 (larger peak current, more slowly decaying tail current). *a-d*, Depolarizations to within 1 mV of -31, -23, -3, and +15 mV. *e, e'*, Families of records taken over the voltage range between +27 and +80 mV to show reversal of Ca channel current with Bay K 8644 (*e'*) and in control (*e*). The current-voltage diagram plots peak current against test potential. Series resistance compensation for 0.75 M Ω out of a total pipette resistance of 0.9 M Ω . Linear leak currents and capacity currents have been subtracted from the current signals. External solution contained (mM) 10 BaCl₂, 85 Cs-aspartate, 2.85 MgCl₂, 5 glucose, and 10 HEPES titrated with CsOH to pH 7.4. Internal dialysate contained (mM) 75.5 Cs-aspartate, 75.5 Cs-phosphate, 5 glucose, and 10 HEPES titrated with CsOH to pH 7.0. Cell 205G.

The Ca agonist altered the voltage- and time-dependence of Ca channel currents carried by either Ca^{2+} or Ba^{2+} . The analysis of Ba current (charge carried by Ba^{2+}) is somewhat simpler because of the absence of Ca-induced Ca channel inactivation²³; Figure 1 illustrates the effect of Bay K 8644 (5 μM) on Ba current in a whole-cell suction pipette recording from a guinea pig myocyte. The agonist-induced enhancement of peak current was largest at weak depolarizations (where the current was small in the absence of drug) and became progressively smaller with increasing depolarization. For example, the enhancement was ~ 10 -fold at -31 mV (*a*) and only ~ 1.8 -fold near $+80$ mV (*e*, *e'*). As the current-voltage relationship indicates, the current through the Ca channel reversed near $+50$ mV (due to a balance of Ba influx and Cs efflux²⁴) and the reversal potential remained unchanged with agonist.

The time-dependence of Ba current through Ca channels was also changed by the Ca agonist. In the presence of Bay K 8644, the current showed a greater tendency to decay. As the depolarizing pulse grew stronger, the enhancement of decay was so great that the agonist current signal was smaller than the control record by the end of the pulse, regardless of the polarity of the current (*d*, *e*, *e'*). Inward tail currents also decayed more slowly following a repolarizing step.

The pronounced changes in kinetics suggest that Ca channel gating is modified by Bay K 8644. To investigate this possibility, we carried out patch clamp recordings of single Ca channel activity.

Long-lasting channel openings

Activity of single Ca channels was recorded from cell-attached patches on guinea pig ventricular cells with 110 mM Ba in the pipette as the charge carrier (Fig. 2*a*, *b*). Ca channel activity was seen in the form of brief pulses of inward current in the absence of drug as in previous studies^{21,25}. The test depolarization was kept small enough to give large, easily resolved unitary currents. Exposure of the cell to Bay K 8644 did not detectably alter the amplitude of the unitary event, but it produced a striking change in the pattern of channel opening. In many of the sweeps, the channel spent the majority of its time open, rather than closed; channel openings were very much prolonged; channel closings were abbreviated. As a result of these effects, the average current signal in the presence of agonist (Fig. 2*d*) was about 11-fold greater than its amplitude before drug (Fig. 2*c*). This response was typical of results from five patches, and reflected the whole-cell response at comparably weak depolarizations (Fig. 1*a*) in the degree of enhancement and in the absence of a drug-promoted relaxation of inward current.

When the agonist effect was large, and the drug dose was at a maximally effective concentration as in Fig. 2*b*, a substantial fraction of sweeps still contained no detectable openings or only rapidly flickering openings like those found in the absence of drug. Such sweeps are easily differentiated from sweeps with long openings and short closings. The combination of both types of behaviour provides an interesting clue to the agonist's mechanism of action.

Sweep-to-sweep variations in behaviour are illustrated for the same experiment by plotting the probability of a channel being open (*p*) for each sweep (Fig. 3*a*, *b*). The value of *p* was obtained by integrating over a period when the current signal was steady (see Fig. 3 legend). In the control run (Fig. 3*a*), *p* was less than 0.20 in all but one of the sweeps. After exposure to Bay K 8644, changes in *p* were immediate and well-maintained throughout the 15 min exposure to the drug (Fig. 3*b*). In the presence of the agonist, the distribution of *p* values is strikingly bimodal: over the range between *p* = 0.5 and *p* = 1.0 a new group of *p* values appears; however, between *p* = 0 and *p* = 0.5, the distribution remains similar to that in the absence of drug (Fig. 3*c*, *d*).

Sweeps with low *p* and high *p*

Taking advantage of the bimodal distribution of *p* values, we separated records taken in the presence of Bay K 8644 into two

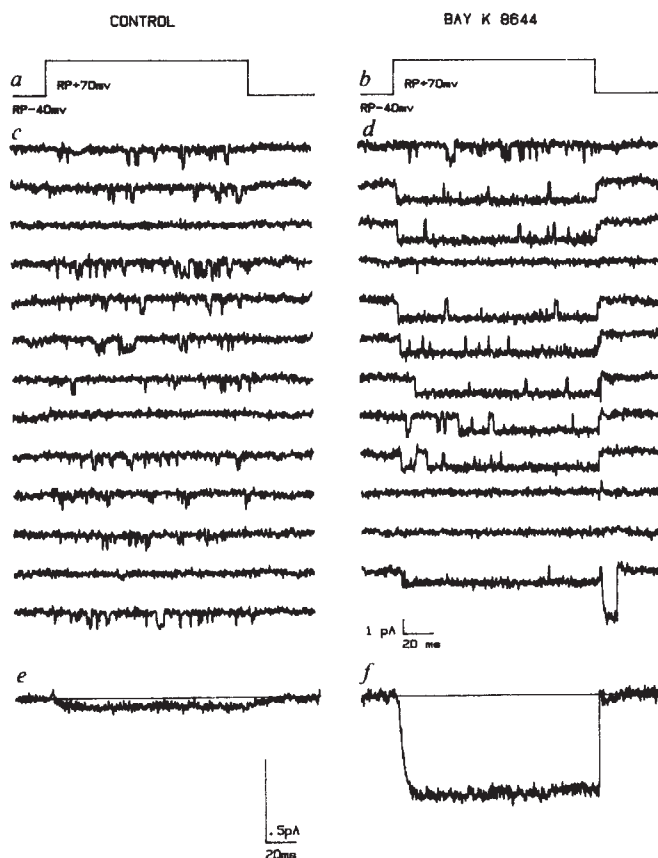


Fig. 2 Patch clamp recording of single Ca channel activity in the absence (left column) and presence of 5 μM Bay K 8644 (right column). Cell-attached patch. *a*, *b*, Voltage protocol RP, resting potential ~ -60 mV as determined by measurement of open channel current-voltage relations (data not shown). Pulses were applied once every 3 s. *c*, *d*, Representative current signals (consecutive sweeps). Linear leak and capacity currents have been subtracted. *e*, *f*, Averaged current signals from all the sweeps in each run (vertical scale is different in *c*, *d*). Cell 6A.

groups of records for analysis of the kinetics of opening and closing: 'low *p*' sweeps with *p* < 0.3, and 'high *p*' sweeps with *p* > 0.65 (Fig. 4). The behaviour of low *p* sweeps in the presence of agonist is very similar to that in the absence of drug (Fig. 4, left and middle columns). Both of these open time histograms are fitted by a single, rapidly decaying exponential ($T \sim 0.5$ – 0.6 ms). Both of the shut time distributions are well-fitted by the sum of two exponentials as described elsewhere^{21,22}. The mean latency-to-first opening is similar in both cases (21 ms control, 19 ms for low *p* sweeps in Bay K 8644). The percentage of sweeps with no openings is also very similar (132/420 = 31% in control and 42/127 = 33% for the low *p* sweeps after agonist).

The behaviour of the high *p* sweeps is strikingly different. The open time histogram shows a predominance of long openings (mean value ~ 20 ms), as expected from the records illustrated in Fig. 3. The distribution fits a single exponential reasonably well, except for a surplus of brief openings in the first few bins, which may reflect imperfect exclusion of the low *p* behaviour with its much greater density of open events per unit time. The shut time distribution is fitted by the sum of two exponentials, as in the other two cases, but the mean closed time is much briefer. Similarly, the cumulative histogram of latency-to-first opening is biased towards shorter times. The combination of longer openings and shorter closings accounts for the extremely high values of *p*.

The results from this patch were representative of a total of three experiments with 5–10 μM Bay K 8644. In an experiment with 100 nM drug, the high *p* sweeps showed similar open and

closed time distributions; the proportion of high p sweeps was simply lower than with maximally effective doses of drug.

Multiple modes of channel gating

The analysis is consistent with the idea that the Ca channel switches between different modes of channel gating. Figure 5a presents our working hypothesis. We use the terms mode 0, mode 1, and mode 2 to refer to the three distinct forms of gating behaviour. Mode 1 is characterized by brief (~ 1 ms) openings occurring in rapid bursts. The most obvious form of gating in the absence of drug, it has been the main focus of single channel recordings up to now^{21,22}. The probability of openness in this mode is relatively low at membrane potentials where the unitary current is easily resolved, although it must be remembered that p can approach levels as high as 0.7 at stronger depolarizations^{21,25}. The observed distributions of open and shut times have been interpreted in terms of two closed states (C1, C2) and one open state (O) (see refs 21, 22, 26). Mode 2 can be distinguished by relatively long-lasting channel openings (~ 20 ms) and typically brief closings. Probability of channel openness is typically high (>0.7) even at relatively negative test potentials. The distributions of open and closed times in Fig. 4, right column are compatible with a C1-C2-O scheme, with very different values of rate constants than those appropriate for mode 1.

Although mode 2 behaviour is most easily studied after administration of Ca agonist, it can also be seen in the absence of drug. For example, Fig. 3a shows a sweep with a p value of ~ 0.85 ; this sweep has openings lasting 3.6, 9.5, 10.8, 3.2, 0.6, and 77.5 ms, typical of what we define as mode 2. The sweep was the only one out of 420 that clearly showed long openings. In a total of 11 patches with one or two channels, sweeps with long openings and unusually high p typically appeared at a rate of about 1%, with a wide range (0.4–11%). Figure 5b–e illustrates the behaviour of the patch with the highest incidence of sweeps showing intrinsic mode 2 behaviour. Two main points can be made. First, sweeps with high p appear significantly clustered in time (Fig. 5b). Second, sweeps with high p in the absence of agonist show a pattern of channel opening and closing (Fig. 5d) very similar to that found in the presence of agonist (Fig. 2b). For example, in the experiment illustrated in Fig. 5 the mean open time in the high p sweeps was 16.4 ms, close to the ~ 20 ms time constant of the slow exponential in Fig. 4f. Similarly good agreement was found in other experiments: average values for the mean open time of high p sweeps were 19.0 in control (93 sweeps, 7 cells) and 18.2 in the presence of Bay K 8644 (338 sweeps, 4 cells).

Mode 0 or null mode refers to a condition or conditions where the channel is unavailable for opening. It is expressed by sweeps showing no openings, in excess of those empty sweeps expected from the C1-C2-O kinetics of mode 1. Voltage-dependent inactivation is one factor that increases the proportion of time a channel spends in mode 0. However, we found an excess of null sweeps even with holding potentials negative enough to achieve maximal removal of voltage-dependent inactivation.

The distinction between mode 0 and 1 can be demonstrated in at least two ways. The cumulative histogram of first latency (Fig. 4d) levels off at about 70%. This runs contrary to a simple C1-C2-O model, which predicts that the saturating level should be 100% (eventually, the channel should reach the open state long enough for an event to be detected). As the curve falls 30% short of certainty it appears that the channel was simply unavailable for opening at the test potential 30% of the time.

Another way of distinguishing mode 0 from mode 1 relies on the histogram of p values, as in Fig. 3c for example. The first two bins of the histogram largely represent sweeps for which no openings could be detected. When these sweeps are disregarded, the remaining distribution rises smoothly to a peak at the fifth bin and then falls smoothly, as might be expected for a single type of gating behaviour, namely mode 1.

By either of these criteria, the vast majority of patches in our experiments show an appreciable number of null sweeps in

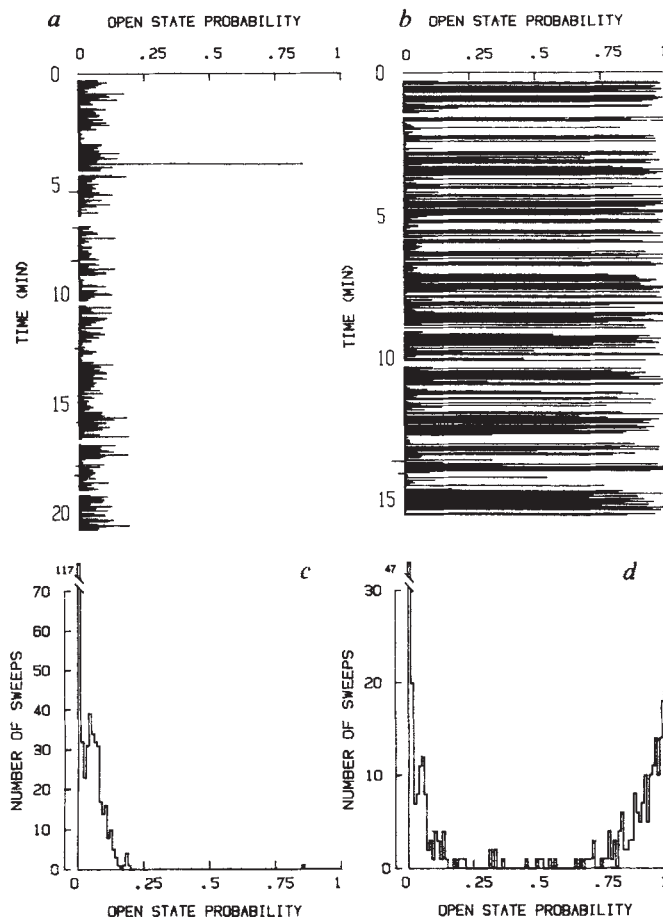


Fig. 3 Influence of Bay K 8644 on probability of channel openness (p) during individual sweeps. Control (left), 5 μ M Bay K 8644 (right). In *a* and *b*, p is plotted against time. Gaps along the time axis indicate sweeps used to measure leak conductance or sweeps marred by noise interference. *c*, *d*, Histograms of p values for the runs shown in *a*, *b*. Each bin is 0.008 wide. For each sweep, steady-state values of p were determined as $p = \langle I \rangle / i$, where i is the unitary current, and $\langle I \rangle$ is the steady-state value of current, averaged over a period from 22 ms after the depolarizing step until the end of the pulse. Sweeps with no detectable openings ('nulls') have near-zero values of p , and show up in the first few bins. For example, in panel *c*, bins 1 to 5 contain 108, 18, 4, 1 and 1 nulls respectively; if the null sweeps had been excluded from the histogram in *c*, the distribution of p values would rise to a single peak and fall smoothly.

excess of those expected from C1-C2-O kinetics. A similar excess has been observed, but not commented on by other investigators (for example, ref. 27). Like sweeps showing mode 2 behaviour, null sweeps show a tendency to follow each other (see Fig. 5 legend).

Mechanism of agonism and antagonism

The identification of different modes of Ca channel gating leads naturally to a hypothesis for the mechanism of action of dihydropyridine (DHP) Ca agonists and antagonists (Fig. 5a). We propose that the binding of these drugs does not plug the pore, but favours one or more modes of gating instead. Within a mode, the kinetics of gating transitions need not be different when DHP occupies its binding site (see Fig. 6). Pure agonist effects arise when mode 2 is specifically favoured, and pure antagonist effects result when mode 0 is preferentially stabilized; these extremes encompass a wide range of intermediate possibilities for 'partial agonist' or 'partial antagonist' effects³⁴.

We observed effects of 5–15 μ M nimodipine consistent with the above hypothesis for a pure antagonist: at these drug concentrations we found that the reduction of the averaged current

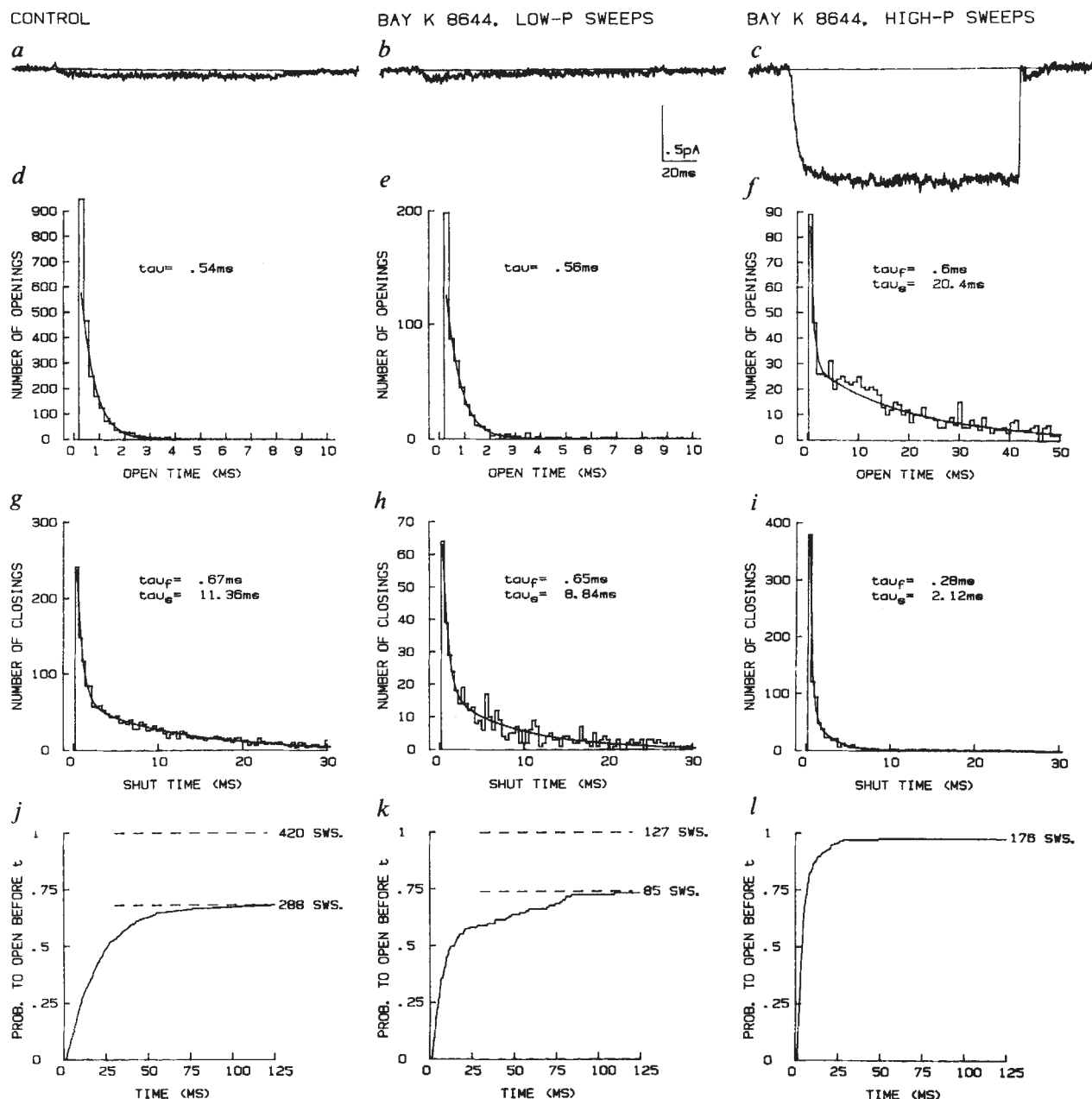


Fig. 4 Effect of Bay K 8644 on Ca channel kinetics. Left column, analysis of all sweeps from the control run; middle column, all sweeps from agonist run with $p < 0.3$; right column, all sweeps from agonist run with $p > 0.65$. The rows show averaged current records (first row), distributions of open times (second row), closed times between openings (third row), and cumulative distributions of the latency to first opening (fourth row). Smooth curves in panels *d* and *e* are single exponentials (the fits ignore the first bin of the histogram data). Smooth curves in panels *f*–*i* are sums of two exponentials. Fits were obtained by minimizing χ^2 , using a computer program provided by Dr R. S. Kass.

was caused by an increase of the proportion of sweeps with no openings, whereas we found no clear evidence for a change in open or closed time distributions.

Mixed effect of nitrendipine

Although nitrendipine is generally assumed to be pure antagonist, we found that the net inhibitory effect of this dihydropyridine is the resultant of agonist as well as antagonist components. Figure 6 illustrates the response of a single Ca channel to 5 μ M nitrendipine, a partially inhibitory dose in the presence of the 110 mM Ba (ref. 24). The control run included a large number of empty sweeps, as well as sweeps with bursts of flickery openings (mode 1), and sweeps with long openings, some lasting beyond the repolarizing step (mode 2, not shown). Nitrendipine altered channel activity in several ways. As with nimodipine, the proportion of sweeps with detectable openings

was greatly decreased. This is evident from the plot of cumulative latencies to first opening (Fig. 6g), which levels off at a level roughly half that of control. When a more complete reduction of inward current was obtained at 10 μ M nitrendipine in other experiments, the proportion of empty sweeps approached 100%. We attribute these effects to stabilization of mode 0, in line with a pure antagonist action.

Other aspects of the nitrendipine response were rather surprising. (1) The percentage of sweeps with $p > 0.5$ was not decreased; (2) a larger percentage of sweeps contained long openings, and the distribution of open times was prolonged (Fig. 6e); (3) closed times were abbreviated (Fig. 6f); (4) latencies-to-first opening were shortened (Fig. 6a). In all of these respects, nitrendipine resembles Bay K 8644.

The agonist-like effect of nitrendipine is most clear at the time of peak inward current. The peak amplitude of the averaged

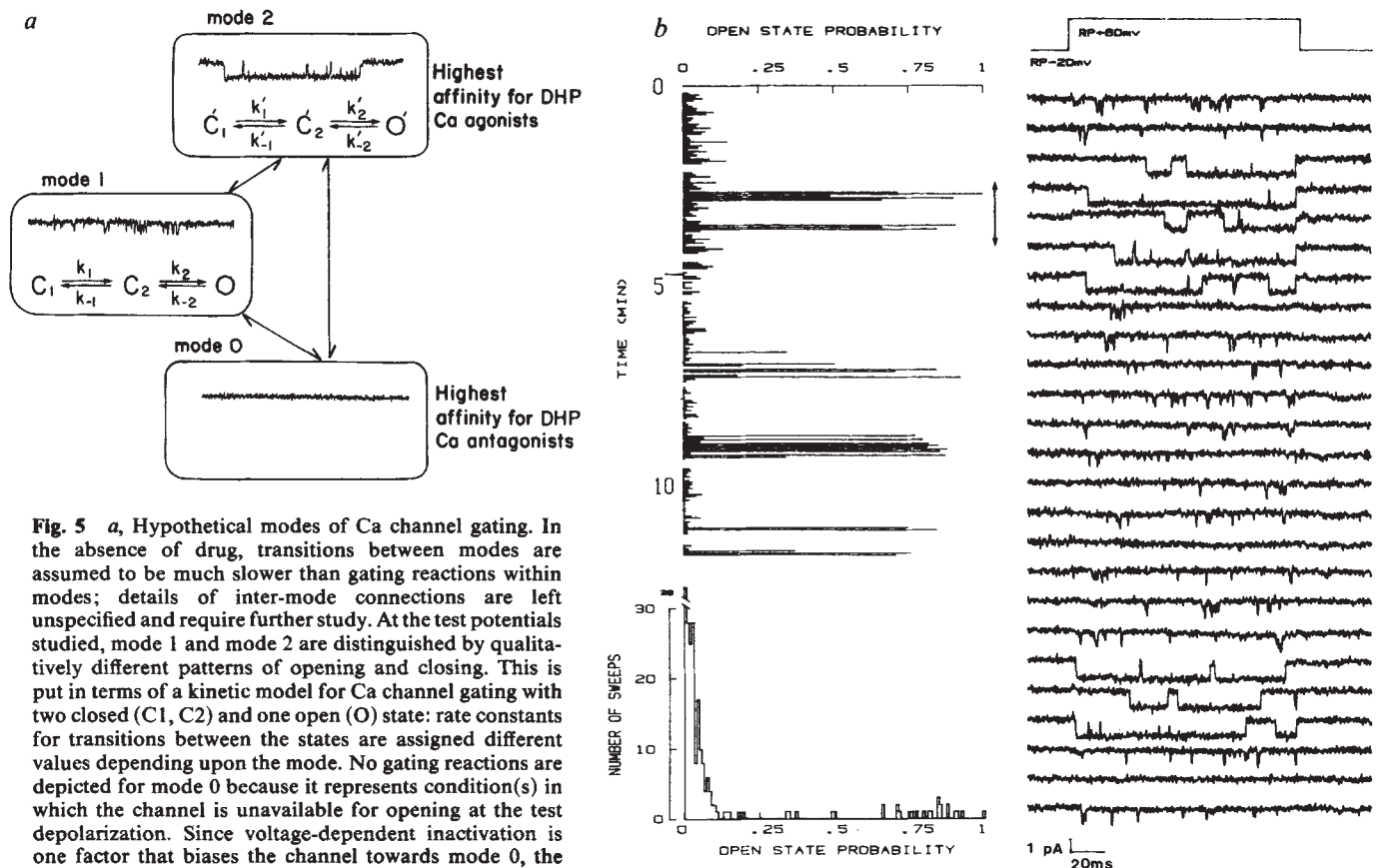


Fig. 5 *a*, Hypothetical modes of Ca channel gating. In the absence of drug, transitions between modes are assumed to be much slower than gating reactions within modes; details of inter-mode connections are left unspecified and require further study. At the test potentials studied, mode 1 and mode 2 are distinguished by qualitatively different patterns of opening and closing. This is put in terms of a kinetic model for Ca channel gating with two closed (C_1 , C_2) and one open (O) state: rate constants for transitions between the states are assigned different values depending upon the mode. No gating reactions are depicted for mode 0 because it represents condition(s) in which the channel is unavailable for opening at the test depolarization. Since voltage-dependent inactivation is one factor that biases the channel towards mode 0, the rate constants for transitions to and from mode 0 are presumably voltage-dependent. The voltage-dependence of transitions between mode 1 and mode 2 is not known. *b-d*, Spontaneous appearance of sweeps with long-lasting openings and high p in the absence of drug. Experimental conditions as in Fig. 2a. RP = -55 mV. *b*, probability of channel openness during 250 sweeps at 0.33 Hz. Gaps correspond to measurements of leak current with weak depolarizing pulses. *c*, Histogram of p values for the set of sweeps shown in *b*. In 31 sweeps, no openings were detected by the analysis program. *d*, Representative current records associated with the depolarizing pulse shown above. Twenty-five records were taken sequentially during the period marked in *b*, illustrating the tendency for mode 2 activity to occur in clusters of consecutive sweeps. The incidence of high p sweeps following immediately after a previous high p sweep was $(11/21) = 52\%$, significantly higher than the overall incidence of high p sweeps $(21/198) = 11\%$ (χ^2 , $P < 0.0001$). The incidence of null sweeps immediately after a preceding null sweep was $(8/29) = 29\%$, significantly higher than the overall incidence of null sweeps, $(30/198) = 15\%$ (χ^2 , $P < 0.05$). Cell 5F. In the presence of Bay K 8644, sweeps with high p values also tended to follow each other. For example, in the run shown in Fig. 3b, the incidence of high p sweeps was $(176/316) = 56\%$ while the incidence of high p sweeps following a previous high p sweep was $(118/176) = 67\%$. This difference is also highly significant, $P < 0.0005$.

current remains almost unchanged (compare Fig. 6c, d), despite a doubling of the number of empty sweeps; this can only mean that p increased in the non-empty sweeps. The increase in p is associated with an increase in the number of long openings relative to the number of short openings (Fig. 6e). The overall inhibitory action of nitrendipine is more obvious toward the end of the depolarization since the average current decays more quickly in the presence of the drug (Fig. 6f). Individual records show a corresponding tendency for openings to occur towards the beginning of the pulse and not towards the end. When the faster decay of mean current was first observed with global current recordings^{24,28}, it was attributed to open channel block²⁴. We now favour the interpretation that the faster decay results from antagonist-promoted inactivation (see below).

Activity of other dihydropyridines

Nitrendipine is not exceptional in its mixed effects. For example, nifedipine is used widely for its inhibitory action, but it can increase cardiac contractility at low doses^{11,29}. The drug CGP 28 392 has agonist effects on contractile activity in heart and smooth muscle¹² and Ca fluxes in cultured cells¹⁶ but its presence partially occludes the stimulatory effect of Bay K 8644 (ref. 16). Bay K 8644 itself is probably not a pure agonist. Sanguinetti and Kass²⁵ have reported that this drug increases Ca channel current in Purkinje fibres when the holding potential is negative, but reduces it when the holding potential is positive to -40 mV.

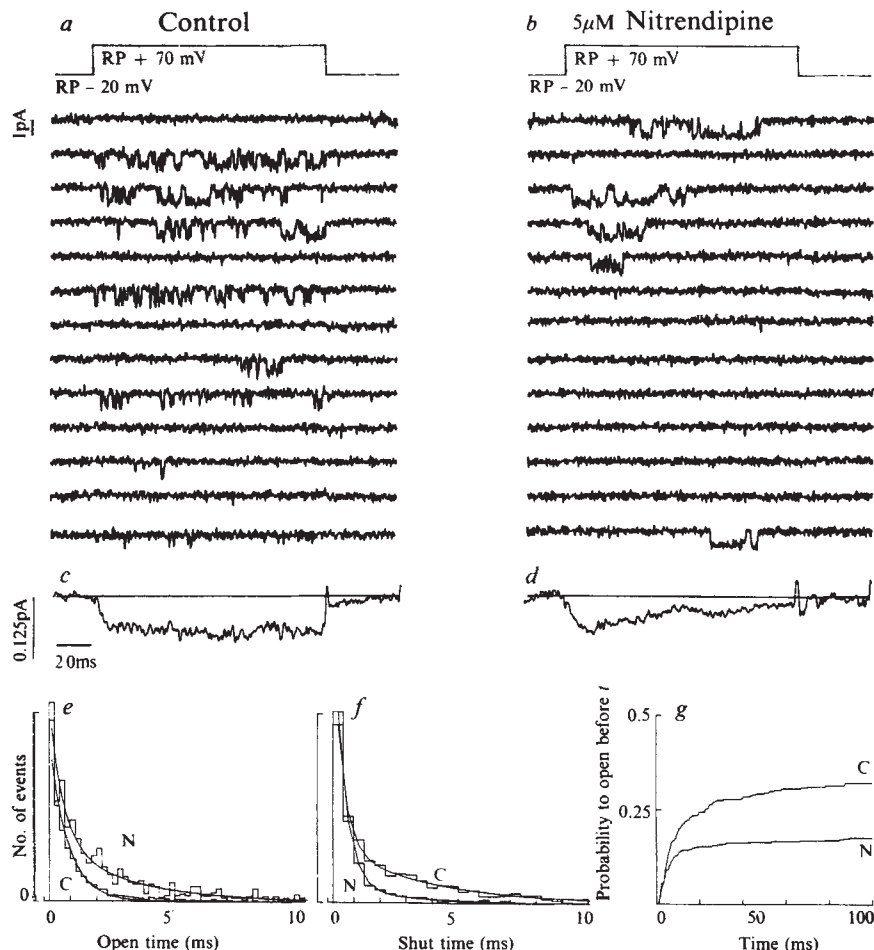
We have seen similar effects of holding potential on the effects of Bay K 8644 in single ventricular cells. A partial antagonist effect may also be involved in the acceleratory effect of Bay K 8644 on Ba current decay with stronger depolarizations (Fig. 1).

Conclusions

Although this work began with an analysis of the effects of Ca agonists and antagonists, the experiments led us to some unexpected conclusions about the fundamental properties of Ca channels. We believe that the strikingly long openings seen with Bay K 8644 are not a new form of gating, but an intrinsic form of gating (mode 2) whose likelihood is merely enhanced by the drug. Without the agonist-promoted activity as a guide, the intrinsic mode 2 activity might be overlooked, since it appears only rarely in most experiments, and contributes only a small number of events per sweep to histograms of open or closed times. But once attention is focused on the mode 2 pattern of long openings and short closings, even single sweeps of this type of activity can be easily picked out. The distinction between mode 1 and mode 2 is particularly clearcut at weak depolarizations where the open probability in mode 1 is low.

Our results provide an attractive explanation for how structurally similar dihydropyridines can exert seemingly opposite effects on Ca influx. They argue against the idea that the inhibitory action of nitrendipine and other Ca antagonists reflects a

Fig. 6 Effect of nitrendipine on the activity of a single channel. Experimental conditions as in Fig. 2a. $RP \approx -60$ mV. **a, b**, Representative groups of consecutive sweeps in the absence (**a**) and presence of $5 \mu\text{M}$ nitrendipine (**b**). Voltage clamp protocol shown above the current traces, **c, d**, Averaged currents from run **a** (409 sweeps) and run **b** (414 sweeps). **e**, Superimposed distributions of open times. Control data (**C**) fitted by a single exponential ($T = 0.8$ ms), nitrendipine data (**N**) fitted by the sum of two exponentials ($T_1 = 0.75$ ms, $T_2 = 3.2$ ms). Full-scale is 700 and 100 openings for control and nitrendipine respectively. **f**, Superimposed distributions of closed times, fitted by sum of two exponentials (control: $T_1 = 0.38$ ms, $T_2 = 3.63$ ms; nitrendipine: $T_1 = 0.42$ ms, $T_2 = 2.57$ ms). Full scale is 600 and 300 closings for control and nitrendipine respectively. **g**, Superimposed cumulative distributions of latencies to first opening. The distribution of open times in nitrendipine (**e**) consisted of a fast exponential component as in control ($T \sim 0.75$ ms), plus an additional slow component. The time constant of the slow component (3.2 ms) falls short of the ~ 20 ms time constants characteristic of mode 2 activity in control and with Bay K 8644. The cause for this discrepancy is not clear. We cannot exclude the possibility that nitrendipine not only influences the proportion of sweeps the channel spends in mode 2, but also affects the kinetics of opening and closing within mode 2. Another explanation is that openings of nitrendipine-bound channels are cut short by an accelerated inactivation reaction (transition from mode 2 to mode 0), which acts in parallel with the closing reaction within mode 2 (O $\rightarrow$ C2) and thereby increases the overall closing rate above its normal value.



physical plugging of the Ca channel pore. A simple binding and plugging hypothesis can be excluded since a nitrendipine molecule cannot plug the channel and show agonist-like activity at the same time; separate binding sites for nitrendipine's antagonist and agonist actions seem highly unlikely since Bay K 8644 competes for all of the nitrendipine binding sites³⁰. The results point instead to modulation of gating as the mechanism of DHP agonist or antagonist action. The changes in gating revealed by the single channel analysis help to explain earlier observations in single cell or multicellular cardiac preparations. All lines of evidence support the idea that inactivation and the inhibitory effects of DHP Ca antagonists reinforce each other. Thus, (1) nitrendipine speeds the decay of Ba current through Ca channels in both whole cell²⁷ and unitary current recordings (Fig. 6), as if drug binding promotes mode 0. (A similar acceleration is seen with Bay K 8644 at stronger test depolarizations (Fig. 1), and is tentatively attributed to the drug's partial antagonist action.) (2) Nitrendipine²⁴ and nisoldipine²⁸ act to slow recovery from inactivation following a depolarizing pulse. (3) Steady depolarizations that inactivate Ca channels strongly promote nisoldipine's inhibitory effect²⁸. (4) Like voltage-dependent inactivation^{21,25}, the inhibitory effect of DHP antagonists is expressed as an increase in the percentage of empty sweeps at the single channel level (Fig. 6).

Our hypothesis provides a simple explanation of partial agonist or partial antagonist effects, and a framework for classifying the 7,000 or more dihydropyridine compounds that remain less well studied than nifedipine or nitrendipine. It will be interesting to see if any of these molecules behaves as a pure agonist or antagonist, exclusively stabilizing mode 2 or mode 0. It is important to remember that the range of possibilities is not limited to a one-dimensional scale between agonism and antagonism, since stabilization of three modes needs to be taken into account. In addition to the ideal agonist and ideal antagonist, one might

imagine a compound that exclusively stabilizes mode 1, and thereby opposes effects of both agonists and antagonists.

The mechanism of transitions between modes is not known. The clustering of mode 2 activity indicates that switching can be relatively slow and raises the question of whether the transition between modes is an expression of a covalent modification of the channel protein. For the moment, we can only say that the modulatory process is probably not cyclic AMP-dependent protein phosphorylation, since the effect of agonist is quite different than that produced by elevation of cyclic AMP with isoprenaline^{25,27} or 8-bromo-cyclic AMP²². Unlike Bay K 8644, isoprenaline slows both the activation and inactivation of Ba currents through Ca channels^{18,27} but it does not by itself slow significantly the decay of inward tails¹⁸; it does not induce mode 2 activity, but it modifies the rate constants within mode 1 (refs 21, 22, 25). Isoprenaline-induced increases in the number of functional channels¹⁸ might be accounted for by a recruitment of channels from mode 0 to mode 1.

Like the cardiac cells described here, sensory neurones from chick dorsal root ganglion and neuroblastoma cells show a mixture of modes 0, 1 and 2, and an enhancement of mode 2 behaviour with Bay K 8644 (refs 31, 32). The apparently widespread existence of the gating modes and their positive response to Ca agonist gives new reason to wonder whether dihydropyridines have endogenous counterpart(s) that up- or down-modulate Ca channels (compare ref. 11).

We thank Drs R. W. Aldrich, C. J. Cohen, A. P. Fox, R. A. Janis, Y. Kurachi, M. C. Nowicky, A. Scriabine and B. Sakmann for helpful discussions, and the USPHS and the Miles Institute for support.

Note added in proof: After submission of this manuscript and publication of preliminary communications of this work^{14,31}, prolonged openings of cardiac Ca channels in the presence of DHP Ca agonists have been reported^{35,36}.

Received 6 March; accepted 10 July 1984

1. Reuter, H. *Nature* **301**, 569-574 (1983).
2. Tsien, R. W. A. *Rev. Physiol.* **45**, 341-358 (1983).
3. Fleckenstein, A. *Calcium Antagonism in Heart and Smooth Muscle* (Wiley, New York, 1983).
4. Flaim, S. F. & Zelis, R. *Calcium Blockers* (Urban and Schwarzenberg, Baltimore, 1982).
5. Schwartz, A. & Taira, N. (eds) *Circ. Res.* **52** (Suppl., part II), 1-181 (1983).
6. Scriabine, A. (ed.) *Nitrendipine* (Urban and Schwarzenberg, Baltimore, 1984).
7. Curtis, B. M. & Catterall, W. A. *J. biol. Chem.* **258**, 7280-7283 (1983).
8. Venter, J. C. *et al. J. biol. Chem.* **258**, 9344-9348 (1983).
9. Norman, R. I., Borsetto, M., Fosset, M. & Lazdunski, M. *Biochem. biophys. Res. Commun.* **111**, 878-883 (1983).
10. Glossmann, H. & Ferry, D. R. *Naunyn-Schmiedeberg's Arch. Pharmac.* **323**, 279-291 (1983).
11. Schramm, M., Thomas, G., Towart, R. & Franckowiak, G. *Nature* **303**, 535-537 (1983).
12. Truog, A. G. *et al. in Calcium in Biological Systems* (eds Rubin, R. P., Weiss, G. B. & Putney, J. W. Jr. (Plenum, New York, 1984).
13. Sanguinetti, M. C. & Kass, R. S. *Biophys. J.* **45**, 394a (1984).
14. Hess, P., Lansman, J. B. & Tsien, R. W. *Biophys. J.* **45**, 394a (1984).
15. Cohen, C. J. & Chung, M. *Biophys. J.* **45**, 394a (1984).
16. Freedman, S. B. & Miller, R. J. *Proc. natn. Acad. Sci., U.S.A.* (in the press).
17. Kao, R. L. *et al. Archs Biochem. Biophys.* **203**, 587-599 (1980).
18. Bean, B. P., Nowicky, M. C. & Tsien, R. W. *Nature* **307**, 371-375 (1984).
19. Lee, K. S., Akaike, N. & Brown, A. M. *J. Neurosci. Meth.* **2**, 51-78 (1980).
20. Hamill, O., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85-100 (1981).
21. Cavalie, A., Ochi, R., Pelzer, D. & Trautwein, W. *Pflügers Arch. ges. Physiol.* **398**, 284-297 (1983).
22. Cachelin, A. B., dePeyer, J. E., Kokubun, S. & Reuter, H. *Nature* **304**, 462-464 (1983).
23. Marban, E. & Tsien, R. W. *Biophys. J.* **33**, 143a (1981).
24. Lee, K. S. & Tsien, R. W. *J. Physiol., Lond.* (in the press).
25. Reuter, H., Stevens, C. F., Tsien, R. W. & Yellen, G. *Nature* **297**, 501-504 (1982).
26. Fenwick, E. M., Marty, A. & Neher, E. *J. Physiol., Lond.* **331**, 599-635 (1982).
27. Brum, G., Osterrieder, W. & Trautwein, W. *Pflügers Arch. ges. Physiol.* **401**, 111-118 (1984).
28. Sanguinetti, M. C. & Kass, R. S. *Circ. Res.* submitted.
29. Himori, N., Ono, H. & Taira, N. *Jap. J. Pharmac.* **26**, 427-435 (1976).
30. Janis, R. A., Sarmiento, J. G., Maurer, S. C., Bolger, G. T. & Triggie, D. J. *J. Pharm. exp. Ther.* (in the press).
31. Fox, A. P., Hess, P., Lansman, J. B., Nowicky, M. C. & Tsien, R. W. *J. Physiol., Lond.* **353**, 75P (1984).
32. Fox, A. P., Nowicky, M. C. & Tsien, R. W. *Soc. Neurosci. Abs.* **10**, 526 (1984).
33. Lee, K. S. & Tsien, R. W. *Nature* **302**, 790-794 (1983).
34. Glossmann, H., Lin, T., Rombusch, M. & Ferry, D. R. *FEBS Lett.* **160**, 226-232 (1983).
35. Ochi, R., Hino, N. & Niimi, Y. *Proc. Jap. Acad. Ser. B* **60**, 153-156 (1984).
36. Kokubun, S. & Reuter, H. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4824-4827 (1984).

LETTERS TO NATURE

Relic interstellar grains in Murchison meteorite

Jongmann Yang & Samuel Epstein

Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA

Isotopic analyses of hydrogen, carbon and nitrogen in meteorites provide important information about the origin and history of these elements both in meteorites and in the Solar System. Here we show that, in the Murchison meteorite, the D/H ratios of hydrogen are unusually high in several separates and in one case up to 30 times the cosmic value of 2×10^{-5} . Many phases show high $^{13}\text{C}/^{12}\text{C}$ ratios, up to 2.5 times the terrestrial value of 0.011. These ^{13}C -rich and D-rich components of the two chemical elements are not correlated. Also they are heterogeneously distributed, suggesting that different components in the meteorite originated from different astrophysical sites and at different times. The D-rich hydrogen in the meteorite is probably due to molecules formed by ion-molecule reactions in interstellar clouds while the tiny amount of ^{13}C -rich carbon is probably due to nucleosynthesis in red giant stars as suggested by Swart *et al.*¹. Both of these heavy-isotope enriched components survived homogenization in the accumulation and subsequent history of the meteorites.

Following studies of isotopic compositions and concentrations of hydrogen and carbon in acid residues from various types of meteorites, we have shown² that: (1) Many primitive carbonaceous and non-carbonaceous meteorites contained HCl-HF insoluble organic matter with high D/H ratios, $> 30 \times 10^{-5}$, suggesting enrichment of deuterium in the molecules by means of isotopic exchange reactions with cosmic hydrogen³⁻⁵ below 150 K. (2) All the meteorites that have been analysed contained inorganic hydrogen which was soluble in HCl-HF solutions. This hydrogen uniformly had a D/H ratio of about 14×10^{-5} , similar to terrestrial hydrogen. This suggests that the D/H ratio in this inorganic component was determined in the solar nebula at a minimum temperature of ~ 200 K by isotopic exchange reactions between molecules and cosmic hydrogen gas. (3) Three unequilibrated meteorites (Semarkona, Bishunpur and Renazzo) showed another extremely anomalous component of acid soluble organic hydrogen with D/H $\approx 170 \times 10^{-5}$, more than 80 times higher than the D/H ratio of cosmic hydrogen. This suggests that meteorites contain organic matter that formed well below 100 K in interstellar clouds before formation of the solar nebula because the temperatures attributed to the solar nebula are much higher than 100 K.

Carbon, on the other hand, showed no obvious ^{13}C anomalies, other than a $\delta^{13}\text{C}$ enrichment of about 60‰ due to the presence

of carbonates^{2,6,7}. δ indicates the variation in isotopic ratios relative to standards in ‰. Standards for hydrogen, carbon and nitrogen are ocean water, PDB and air, respectively. However, a tiny fraction of the carbon, was found to be resistant to various chemical reagents and resistant to combustion below 800 °C. This carbon was a new component in the meteorite and its $^{13}\text{C}/^{12}\text{C}$ ratio was greater than all carbon previously found on the Earth and in the meteorites^{1,8,9}. This carbon has a $^{13}\text{C}/^{12}\text{C}$ ratio of at least twice that of the Earth.

The aim of the present study was to determine how the new ^{13}C -rich carbon might relate in origin to the D-rich hydrogen and to the $^{15}\text{N}/^{14}\text{N}$ ratio of nitrogen. Murchison meteorite was powdered and aliquots of the samples were demineralized by reaction with HCl-HF solutions. The resulting residue, designated CF, was then treated with various chemicals to remove certain components of the residue and separated by particle size (see Table 1) to try to concentrate the exotic mineral phases and the interstellar relic grains. Gas samples of hydrogen, carbon and nitrogen were analysed by standard stable isotope ratio mass spectrometric techniques and reported in the usual δ notations.

Concentrations and isotopic data of hydrogen, carbon and nitrogen are given in Table 1. For the bulk sample the hydrogen is dominated by hydrogen from inorganic sources whose δD value is about -50‰. The $\delta^{13}\text{C}$ value of the total carbon is about -2.4‰ but the presence of heavier carbon is apparent from the CO_2 given off at the combustion above 500 °C ($\delta^{13}\text{C} = 27\%$). The D/H data of the CF residue show that all inorganic hydrogen is removed by HCl-HF acid and a deuterium-rich organic material dominates the residue (Fig. 1a).

Two residue samples (CFOc, CFOf) resulting from the oxidation of CF residue by reagent H_2O_2 contain an unusually high $^{13}\text{C}/^{12}\text{C}$ ratio carbon which is oxidized to CO_2 above 800 °C. The $\delta^{13}\text{C}$ ranges between 500 and 1,200‰. In principle, such high $\delta^{13}\text{C}$ values can be produced by isotope exchange reactions at very low temperatures¹⁰ that exist in the interstellar media. In such circumstances, one might expect that there would be a much greater heterogeneity in the $\delta^{13}\text{C}$ values in the organic residues and that a much larger fraction of the organic acid residue would be relatively enriched in ^{13}C . Therefore, it is the tiny amount of ^{13}C -rich component that makes it a good candidate for a red giant origin. On the other hand, large fractions of the organic residues have high δD values. As we have argued previously, such δD values for a large fraction of the organic residues could only be obtained by ion-molecule reactions occurring in the interstellar medium.

The finer grain fraction (CFOf) has a hydrogen and carbon content per gram of bulk meteorite a factor of 10 times higher than the coarser grain fraction (CFOc). Compared with CFOc, CFOf has a higher C/H ratio of 2.76 and a higher $^{13}\text{C}/^{12}\text{C}$ ratio of up to 0.025 but a lower D/H ratio by a factor of two. Perhaps

Table 1 The concentrations and isotopic compositions of hydrogen, carbon and nitrogen released by stepwise oxidation-pyrolysis of bulk and various acid residues of Murchison meteorite

Sample*	T† (°C)	H ₂ ($\mu\text{mol g}^{-1}\ddagger$)	δD	CO ₂ ($\mu\text{mol g}^{-1}\ddagger$)	$\delta^{13}\text{C}$	N ₂ ($\mu\text{mol g}^{-1}\ddagger$)	$\delta^{15}\text{N}$	C/H§
Bulk	300	1,738	-48	310	1.6			0.089
51.9 mg	350	454	-47	250	-3.9			0.28
	500	820	-51	840	-7.9			0.51
	900	217	-80	120	27.0			0.29
	<u>1,100</u>	~20	-48	~1	27.6			—
Total		3,250	-51	1,530	-2.4			0.23
CF	300	47.9	463	42	-3.9			0.44
4.79%	350	10.0	562	68	-10.4			3.39
20.5 mg	500	1.5	607	129	-20.7			42
	800	<0.1	—	7.3	-8.3			>38
	<u>1,100</u>	<0.1	—	~0.5	-22			—
Total		59.4	483	246	-14.5			2.07
CFOc	(250)	1.49	1,042	0.71	-9.5			0.24
0.39%	(300)	0.73	1,104	0.33				0.22
20.1 mg	(350)	0.74	1,708	0.65	-11.9			0.44
	700	5.97	2,584	33.5	-6.5	0.23	-13.2	2.81
	800	0.49	2,213	0.23	176	<0.02	—	0.22
	900	~0.09	~2,853	0.11	508			0.47
	<u>1,100</u>	~0.04	—	0.77	184			-6.1
Total		9.56	2,144	36.3	0.2			1.90
CFOf	(250)	16.7	268	11.2	-19.2			0.34
1.18%	(300)	6.0	—	5.0	-19.3			0.42
35.1 mg	(350)	6.9	791	6.2	-20.9			0.45
	350	19.5	992	17	-20.9			0.43
	500	28.2	1,008	76	-21.9			1.35
	700	12.6	888	98	-22.0			3.90
	800	7.9	823	71	-22.0	1.5	-20.8	4.51
	900	4.1	826	68	-22.6	1.2	-28.7	8.24
	800	1.9	930	221	-21.9	1.3	-21.6	58
	900	0.09	1,111	0.04	~703	<0.02	—	-0.2
	<u>1,100</u>	<0.006	—	0.05	~1,220	<0.02	—	—
Total		104	806	574	-21.7			2.76
CFOfN(I)	(250)	0.56	-45	0.65	-23.6	<0.01	—	0.58
0.12%	(350)	0.26	41	0.73	-24.1	<0.01	—	1.38
7.2 mg	500	3.25	309	46.2	-30.6	0.19	-170	7.12
	700	0.18	158	3.9	22.6	0.02	33.5	10.7
	800	<0.01	~600	0.09	985	<0.01	—	—
	<u>1,100</u>	<0.01	~700	0.05	~1,480	<0.01	—	—
Total		4.25	240	51.6	-23.2			6.03
CFOfN(II)	(350)	0.51	-158	1.02	-25.9	<0.06	—	1.0
0.12% 0.6 mg	700	1.31	57	42.7	-22.7	0.23	—	16
CFOfNf	(350)	3.5	124	5.4	-25.1	<0.01	—	0.78
~0.17%	500	5.1	158	44.0	-25.0	0.17	6.7	4.36
5.0 mg	700	2.3	34	0.43	-20.6	<0.01	—	0.095
	800	<0.1	—	0.04	-24.5	<0.01	—	—
	<u>1,100</u>	<0.1	—	0.03	-25.5	<0.01	—	—
Total		10.8	121	49.9	-25.0			2.31
CFP	(350)	0.10	23	0.09	—	<0.01	—	0.45
0.24%	500	0.24	214	6.72	-36.7	<0.01	—	14
11.0 mg	700	<0.1	—	1.79	-28.6	<0.01	—	—
	<u>1,100</u>	<0.1	—	0.16	977	<0.01	—	—
Total		0.34	158	8.76	-16.3			13
P	(350)	150	-81	49	-28.4	0.7	—	0.16
34.3%	500	210	-79	20	-36.4	<0.1	—	0.049
129.8 mg	800	68	-101	0.40	81	0.14	—	0.003
	<u>1,100</u>	13	-117	~0.35	~573	0.1	—	0.013
Total		440	-84	70	-27.1			0.080
PCF	(350)	0.44	~144	0.51	-22.6	<0.05	—	0.58
2.66%	700	<0.02	—	4.39	-31.7	~0.03	—	>60
8.2 mg	<u>1,100</u>	<0.02	>300	~0.22	~705	~0.03	—	—
Total		0.44	~144	5.12	~1.1			5.8
CFS	300	101	691	99	-7.1			0.49
2.2%	800	76	785	450	-17.0			2.94
5.6 mg	900	4.8	659	0.44	320			0.05
	<u>1,100</u>	2.7	~594	~0.09	~630			—
Total		185	728	550	-14.8			11.48

Eight acid treated residues (from CF to PCF) were prepared from aliquots of Murchison bulk powder. The procedures were basically similar to previous studies^{2,12}. An aliquot of HCl-HF treated residue (CF) which contains sulphides, spinel, chromite and organic polymer was further treated with acetone, hexanes, CS₂, benzene, methanol, 6M NaOH and 30% H₂O₂. This residue, CFO, was separated into a very fine-grained colloidal phase in an H₂O solution (CFOf) and into a coarse-grained residue (CFOc). CFOf was black, its density was 1.25 and it contained the organic polymer, sulphides and spinel. Residue CFOc was grey and contained spinel, chromite and organic polymer. An aliquot of CFOf was treated with boiling NaOH-NaOCl and fuming HNO₃. Two new residues were produced. One (CFOfN) contained organic polymer, chromite and spinel settled out on centrifugation, and the other (CFOfNf) remained dispersed in the solution of HCl-acetone but was recovered by evaporation to dryness. It contained organic polymer, sulphides and spinel. An aliquot of CF was boiled in HClO₄ producing a residue, CFP which was white and contained spinel and chromite. To examine the chemical resistivity of trace quantities of ¹³C-rich carbon present in Murchison, an aliquot of bulk Murchison was dissolved with boiling 70% HClO₄ producing a residue, P which was white and contained silicates and spinel. An aliquot of this residue P was further treated with HCl-HF. This residue PCF contained spinel and silicates. Residue CFS² which contained organic polymer, chromite, sulphides and spinel was prepared treating bulk Murchison with HCl-HF solution and with CS₂.

* Acid residues: C, HCl; F, HF; O, H₂O₂; N, HNO₃; P, HClO₄; S, CS₂; f, finer grain fraction; c, coarser grain fraction. Weight percentage of the acid residue and amount of sample used for the analysis are also shown.

† Samples were heated in a quartz vacuum system for 1 h at each temperature in an oxygen atmosphere of ~5 mmHg except the underlined steps where the samples were heated for 4–18 h. T in parentheses represents the step where pyrolysis was made in the absence of oxygen. The pyrolyses steps were to analyse separately the terrestrial contamination.

‡ Yields are given in $\mu\text{mol g}^{-1}$ of bulk Murchison.

§ Very low C/H ratios are mainly due to dominance of inorganic hydrogen and very high C/H ratios are due to carbon accumulated at lower temperature steps. Thus each C/H ratio at each T step should not be interpreted as a characteristic average C/H ratio of the organic matter decomposed at that temperature.

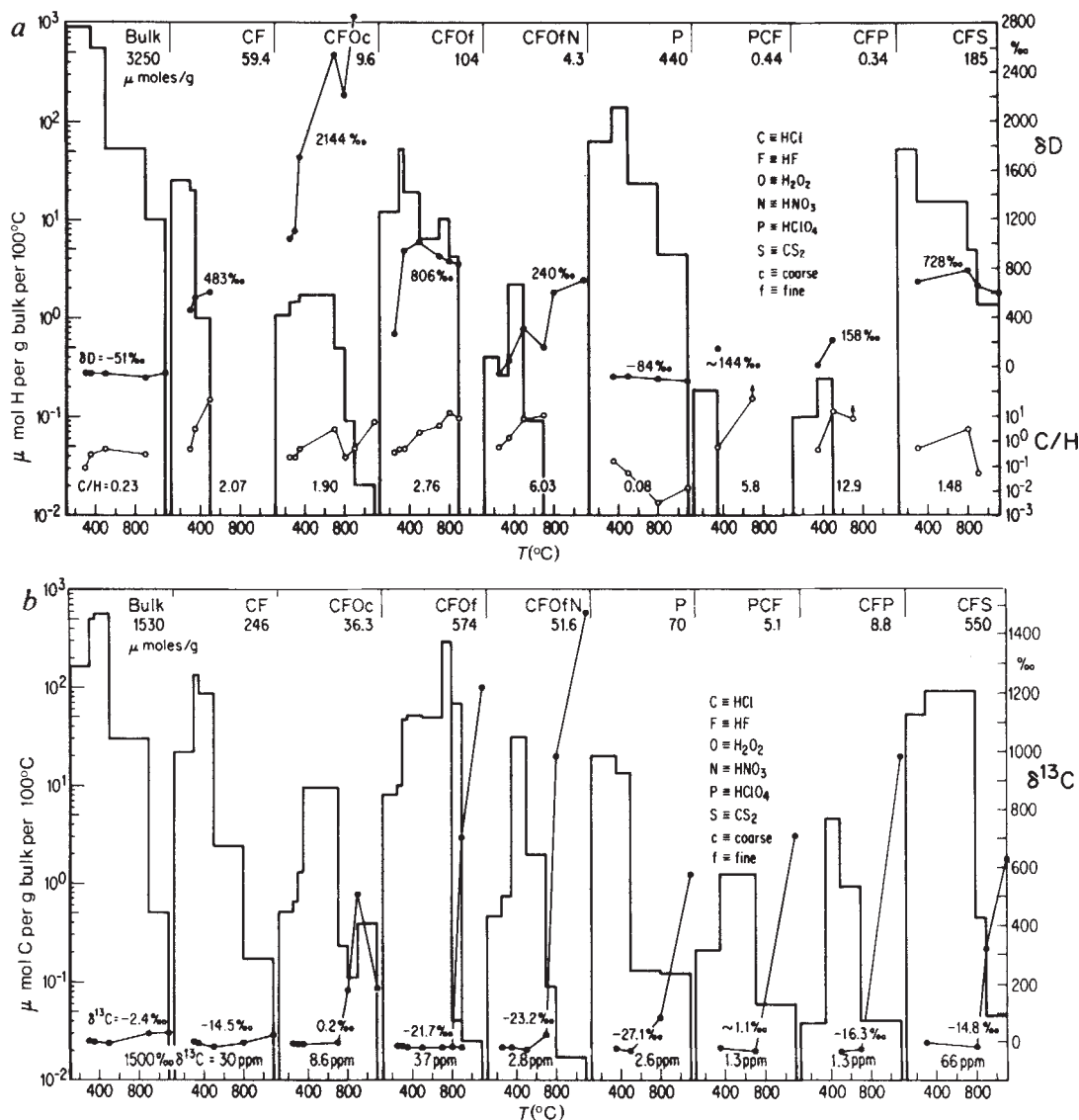


Fig. 1 a, A histogram showing the amount of hydrogen released by stepwise oxidations (pyrolysis for lower temperature steps up to 350 °C) of Murchison bulk and acid treated residue samples plotted against the releasing temperature. δD values (●) and C/H ratios (○) of the released gases are also shown. A sample (CFOc) which is composed primarily of spinel and chromite shows the highest δD values of 1,040–2,850‰. Perchloric acid treated samples (P, PCF, CFP) show much lower δD values. Also included are: values for the total H_2 (in $\mu mol g^{-1}$), its δD value and the C/H ratio of the total gases. **b,** The carbon released simultaneously with the hydrogen. The concentration of the total carbon, the $\delta^{13}C$ value of the total carbon and the amount of carbon which has $\delta^{13}C$ value of 1,500‰ in units of $\mu g C$ per g of bulk sample are given.

this means that the CFOc residue, 60% of which is spinel, contains a larger fraction of the preserved interstellar organic matter. The CFOf residue, 70% of which is organic polymer, perhaps contains relatively more of the organic matter formed in the solar nebula at higher temperatures and thus has a lower D/H ratio. Note that HCl–HF residues of many meteorites that we previously examined² also contained large amounts of spinel and chromite.

A residue sample (CFOfN) formed by treating CFOf with boiling NaOH–NaOCl solution and fuming HNO_3 shows extremely high $\delta^{13}C$, up to ~1,480‰, and a C/H ratio of 6.0 (Table 1). If it can be assumed that the total carbon in each residue is composed of two carbon sources with nominal $\delta^{13}C$ values of 1,500‰ and –30‰, then residues CF, CFOc, CFOf and CFOfN contain respectively 30, 9, 37 and 3 p.p.m. of the isotopically-heavy carbon component (Fig. 1b).

The concentration and $\delta^{15}N$ of nitrogen was also measured for residue samples (Table 1). In the case of the CFOfN sample, most of the nitrogen was released at the 500 °C combustion temperature and the $\delta^{15}N$ value for this N_2 was –170‰, lower than for any other fraction. Our data for total nitrogen ($0.21 \mu mol g^{-1}$, $\delta^{15}N = -151‰$) for this sample checks well with the value of Lewis *et al.*¹¹ ($0.16 \mu mol g^{-1}$, $\delta^{15}N = -205‰$). There is no correlation between this light nitrogen and isotopically-heavy carbon. Lewis *et al.* came to the same conclusion from their study of acid residues from Allende and Murchison meteorites. They suggested that light nitrogen was produced by nucleosynthesis which is responsible for CCF–xenon. We now

suggest an alternative explanation for the –170‰ nitrogen. There seems to be some relationship between the δD and $\delta^{15}N$ in our residues. Thus, the isotopic composition of the nitrogen and the deuterium may be attributable to ion–molecule reactions in the interstellar medium.

Another technique was used to try to isolate the ^{13}C -rich carbon in the residues. The CF residue was boiled in 70% $HClO_4$ and labelled as CFP. The CFP sample contained a concentration of $0.16 \mu mol g^{-1}$ of carbon whose $\delta^{13}C$ is 977‰. These are comparable with that observed for residues CFOf and CFOfN (Table 1). However, the CFP residue contained much less total hydrogen and total carbon and almost no nitrogen compared with residues CFOf and CFOfN. Thus perchloric acid is very effective in removing most of organic material but not the very heavy carbon. We, therefore, conclude that the ~1,500‰ $\delta^{13}C$ carbon is not associated with hydrogen and nitrogen. It is quite possible that this small carbon fraction is in a carbide form such as Fe_3C and SiC , or is present as some high-temperature carbon grain such as diamond. Also this result suggests that the low ^{15}N nitrogen and the ^{13}C -rich carbon do not come from the same source.

The perchloric treatment was repeated on an aliquot of Murchison bulk sample. This procedure destroyed 95% of the carbon but the remaining residue (P) contained a heavy carbon component whose $\delta^{13}C$ is ~570‰. This sample, however, contains a large amount of inorganic hydrogen with a characteristic δD value of –80‰, confirming that the hydrogen in this sample is dominantly inorganic. The perchloric acid treatments (P, PCF,

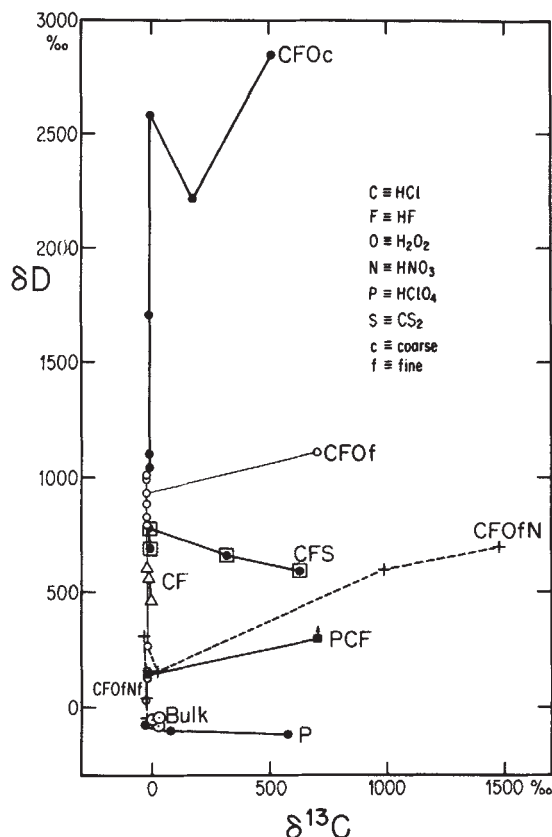


Fig. 2 A plot of the δD against the $\delta^{13}C$ of hydrogen and carbon released by oxidation of Murchison samples. Carbon with high $\delta^{13}C$, up to 1,480‰, was found in the CFOfN residue. This residue consists mostly of carbon with some chromite, spinel and sulphides. There is no clear correlation between the high δD hydrogen and the high $\delta^{13}C$ carbon in the various release patterns and in the various residues, suggesting that they are located in different minerals or different grains. Note that all samples except bulk, CF and CFOfN have a high $\delta^{13}C$ component which is released on oxidation at high temperatures. The carbon with high $\delta^{13}C$ values compared with the total carbon is preferentially resistant to all chemicals we have used.

CFP) show that the ^{13}C -rich carbon comes from a different source to the deuterium.

We have compared the $\delta^{13}C$ values with δD values of coexisting carbon and hydrogen extracted from various components. This is shown in Fig. 2. The most striking feature of this plot is that the $\delta^{13}C$ values initially recorded up to temperatures of 800 °C are on average about -20‰ whereas the δD values of the hydrogen cover an enormous range of -80 to +2,600‰. On the other hand, at the higher temperature range of extraction of the carbon where $\delta^{13}C$ values range up to 1,500‰, the accompanying hydrogen has relatively smaller up and down variations in δD values. It is difficult to assign a common chemical origin of the meteorite to the D-rich hydrogen and the ^{13}C -rich carbon. The previously mentioned perchloric acid experiments suggested a similar conclusion.

Our experimental procedures will probably add contaminants of low values of δD and $\delta^{13}C$. Consequently, our high $\delta^{13}C$ and δD values are probably minimum values and we hope that procedures will eventually be perfected to allow measurement of the pristine δD and $\delta^{13}C$ values. In the case of hydrogen, the higher the δD values for the extracted hydrogen, the more confident we can become about the proposal that the organic matter and its isotopic composition originated from an interstellar medium due to ion-molecule hydrogen isotopic exchange reactions.

Thus, unusually high $\delta^{13}C$ carbon can be isolated by treating the Murchison meteorite with certain reagents due to the preferential destruction of the ^{12}C -rich components. The fine-grained

acid residue (CFOf) contains the highest concentration of the ^{13}C -rich carbon. The ^{13}C -rich component has special chemical properties which will allow its isolation and identification.

The δD of hydrogen does not correlate with the $\delta^{13}C$ of carbon. The coarse-grained acid residue (CFOc) contains hydrogen of the highest D/H ratio while the $\delta^{13}C$ enrichment is relatively modest. The acid residue (CFOfN) which contains the highest ^{13}C -rich component has only modest D enrichment. If the D-rich component is formed by ion-molecule isotope exchange reactions in the interstellar dust clouds, the ^{13}C -rich component is not associated with the deuterium enrichment. This adds credibility to the suggestion by Swart *et al.*¹ regarding the red giant origin of the ^{13}C -rich component in the meteorite.

The fact that the ^{13}C -rich component survived the post accumulation history of Murchison without losing its high ^{13}C value supports our previous assertion² that the $\delta^{13}C$ of the major part of the organic carbon in the meteorite is representative of the carbon in the interstellar medium. In addition, the presolar origin of the ^{13}C -rich component in the acid residues of the meteorites agrees with the suggestion that the high D/H ratios observed in the organic matter in the meteorites were due to the preservation of the presolar material. After the D-rich organic matter was incorporated into the meteorite the post-accumulation history for the organic matter should have been sufficiently mild to allow its preservation and its availability for our examination.

We thank Dr I. R. Kaplan for nitrogen analyses, J. Ruth for technical assistance, and Dr E. Olson for a specimen of Murchison meteorite. This research was funded by NASA (NGL-05-002-190) and by NSF (EAR-7816873).

Received 16 May; accepted 24 July 1984.

- Swart, P. K., Grady, M. M., Pillinger, C. T., Lewis, R. S. & Anders, E. *Science* **220**, 406-410 (1983).
- Yang, J. & Epstein, S. *Geochim. cosmochim. Acta* **47**, 2199-2216 (1983).
- Watson, W. D. *Rev. Mod. Phys.* **48**, 513-552 (1976).
- Geiss, J. & Reeves, H. R. *Astr. Astrophys. J.* **93**, 189-199 (1981).
- Robert, F. & Epstein, S. *Geochim. cosmochim. Acta* **46**, 81-95 (1982).
- Clayton, R. N. *Science* **140**, 192-193 (1963).
- Smith, J. W. & Kaplan, I. R. *Science* **167**, 1367-1370 (1970).
- Kerridge, J. F. *Lunar planet. Sci. Conf.* **13**, 381-382 (1982).
- Yang, J. & Epstein, S. *Meteoritics* **18**, 429-430 (1983).
- Langer, W. D., Graedel, T. E., Frerking, M. A. & Armentrout, P. B. *Astrophys. J.* **277**, 581-604 (1984).
- Lewis, R. S., Anders, E., Wright, I. P., Norris, S. J. & Pillinger, C. T. *Nature* **305**, 767-771 (1983).
- Alaerts, L., Lewis, R. S. & Anders, E. *Geochim. cosmochim. Acta* **43**, 1399-1415 (1979).

Improved positive-ion composition measurements in the upper troposphere and lower stratosphere and the detection of acetone

G. Hauck & F. Arnold

Max-Planck-Institut für Kernphysik, PO Box 103 980, D-6900 Heidelberg, FRG

Atmospheric ion composition measurements were recently extended downwards to the lower stratosphere and upper troposphere following development of a new aircraft-borne mass spectrometer technique. Perhaps the most interesting aspect of these measurements is that they can serve as a powerful tool for neutral trace gas detection in passive chemical ionisation mass spectrometry (PACIMS). However, the measurements, in particular those of positive ions, suffer from a poor mass resolution which complicates ion identification and the PACIMS application. Here we report improved positive-ion composition measurements in the lower stratosphere and upper troposphere, using an improved aircraft-borne mass spectrometer with higher mass resolution and sensitivity than hitherto achieved.

The experimental method is essentially similar to one used for our previous measurements^{1,2}, but with four modifications:

Table 1 Positive ion masses obtained from high-resolution spectra taken in flights F4 and F5 at 11,280 m altitude

Mass	Ion
28 ± 1	HCO ⁺
	H ⁺ HCN
32 ± 1	H ⁺ CH ₃ OH
37 ± 1	H ⁺ (H ₂ O) ₂
50 ± 1	H ⁺ CH ₃ OH·H ₂ O
55 ± 1	H ⁺ (H ₂ O) ₃
59 ± 1	H ⁺ (CH ₃) ₂ CO
64 ± 1	H ⁺ CH ₃ COH·H ₂ O
	H ⁺ C ₂ H ₅ OH·H ₂ O
	H ⁺ CH ₃ OCH ₃ ·H ₂ O
	H ⁺ HCOOH·H ₂ O
73 ± 1	H ⁺ (H ₂ O) ₄
77 ± 1	H ⁺ (CH ₃) ₂ CO·H ₂ O
88 ± 1	H ⁺ CH ₃ OH(H ₂ O) ₂
114 ± 1	H ⁺ (CH ₃) ₂ CO(H ₂ O) ₃
126 ± 1	H ⁺ (H ₂ O) ₇

Tentative ion identifications are also given. Instead of H₂O, ions may contain an ammonia molecule although this seems to be less likely.

(1) a larger quadrupole mass spectrometer (rod diameter, 0.8 cm; rod length, 20 cm) providing higher mass resolution and ion transmission; (2) on-axis ion sampling providing a higher sampling efficiency; (3) an ion-dissociation mode allowing for fragment ion studies; (4) a high-frequency glow discharge ion source working at ambient gas pressure and allowing for in-flight calibration and testing.

The modified instrument will be described in detail elsewhere together with laboratory simulation studies; here the ion-dissociation mode will be briefly discussed. After passing through the mass spectrometer inlet hole (diameter, 0.01 cm), located at the tip of a conically-shaped sampling electrode onto which the incoming gas stream is directed, atmospheric ions are electrically extracted from the high-density gas jet forming behind the inlet hole. During extraction they undergo energetic collisions with gas molecules stimulated by the electric field established between the four-pole rods and the sampling cone. As a result, efficient collisional breakup of the ambient cluster ions occurs and thus only fragments, naked- or weakly-clustered 'cores', of the ambient complex cluster ions are ultimately mass analysed and detected.

Typically, the sampling cone is kept at ground potential and the four-pole field axis is several tens of volts, depending on mass resolution. This ion dissociation mode offers two advantages, a larger sensitivity and a better 'core ion' identification. Upper tropospheric positive ions are expected² to be heavily hydrated (~6–7 H₂O-ligands) and, therefore, very massive. Fragmentation leads to much smaller ions, which are more successfully transmitted through the four-pole rods (particularly at high mass resolution) and easier to identify.

A few important characteristics of the new instruments are: mass range, 0–1,000 AMU; smallest mass peak width at half peak amplitude $\Delta m = 1$ AMU; sensitivity, ~0.1 ion counts per ion cm⁻³ (depending on mass resolution).

The new instrument has been flown twice (flights F4 and F5) on board a twin jet research aircraft (type FALCON 20) of the DFVLR (Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt) over the FRG. The flights took place during daytime on 25 November 1983 (F4) and 24 February 1984 (F5). Maximum altitudes of 11,280 m were reached in both flights while the local tropopause heights were 11,900 m (F4) and 9,750–10,036 m (F5). High mass-resolution positive-ion data were obtained only at the float height (11,280 m) where sufficiently long integration times were available and they thus relate to the uppermost part of the troposphere (F4) and the lowermost part of the stratosphere (F5). Lower mass-resolution data obtained in the height region down to 5,300 m will not be discussed here.

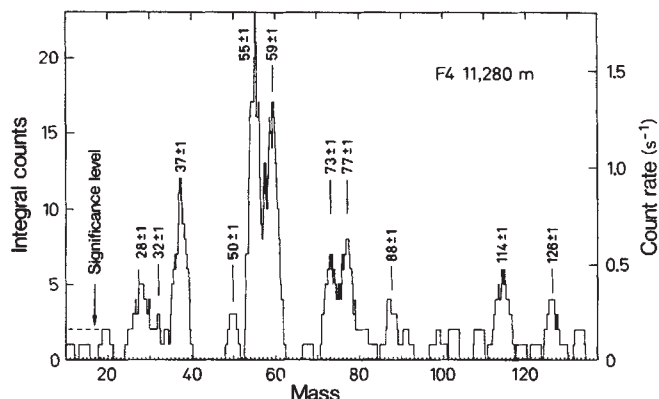


Fig. 1 Positive ion mass spectrum as obtained in flight F4 at 11,280 m altitude. A 2σ significance level is indicated. Total integration time, 640 s.

A typical positive-ion mass spectrum ($\Delta m = 3$ AMU) as measured in F4 at 11,280 m is shown in Fig. 1. As confirmed by the larger mass range, but lower mass-resolution studies, made in F4 at the same height, all major positive ion peaks fall into the mass range shown by the spectrum. Among these, 55 ± 1 and 59 ± 1 are most prominent peaks followed by 37 ± 1 , 73 ± 1 , and 77 ± 1 . A compilation is given in Table 1 along with tentative ion identification. The ions 37 ± 1 , 55 ± 1 , and 73 ± 1 may readily be identified as proton hydrates ($H^+(H_2O)_n$; $n = 2, 3, 4$) which were previously found in the stratosphere^{3–5} and which are expected^{6,7} to form quickly at the tropopause from primary positive ions produced by galactic cosmic-ray ionization. In view of the mass uncertainty of ± 1 AMU, which is mainly due to poor ion counting statistics, it is also possible that the core ion is NH_4^+ instead of H_2O .

The size distribution of the observed $H_3O^+(H_2O)_n$ or $NH_4^+(H_2O)_n$ clusters peaks at n equal to 2 which is quite different from the distribution of ambient $H_3O^+(H_2O)_n$ ions² which should peak at around n equal to 6 and 7. Thus, there is clear evidence for massive electric field-induced dissociation which greatly facilitates the identification of other observed ions.

To improve mass identification of the most prominent peaks observed in F4 (55 ± 1 and 59 ± 1), a much higher mass resolution ($\Delta m = 1$ AMU) was used in F5. However, most of the mass range had to be sacrificed for the high-resolution study to gain integration time for the mass range of interest (50–71 AMU), and thereby to offset the lower sensitivity associated with higher resolution. The portion of the spectrum scanned in the high-resolution mode ($\Delta m = 1$ AMU) of F5 is shown in Fig. 2 along with a lower resolution ($\Delta m = 3$ AMU) spectrum of the same flight. Surprisingly, the spectrum is quite different from that shown in Fig. 1 with 55 ± 1 being virtually absent. However, peak 59 is present, and its mass can be unambiguously identified. The large mass range, low-resolution studies made in F5 reveal that peak 59 is the most prominent one in the entire mass range, and that only minor peaks of other ions are present.

To identify ion 59, the following arguments should be considered: (1) the ion seems to be an unhydrated 'core ion' as no fragment having a mass of 41 is observed; (2) the hydrate of 59 seems to have a stability which is similar to the stability of $H_3O^+(H_2O)_3$ or $NH_4^+(H_2O)_3$ as suggested by the similarity of the ratios $77 \pm 1/59$ and $73 \pm 1/55 \pm 1$; (3) ion 59 is probably a fragment of ions formed by the reaction of $H_3O^+(H_2O)_n$ with an atmospheric trace gas which possesses a proton affinity sufficiently larger than that of H_2O (165.3 kcal mol⁻¹)⁸.

If correct, the first argument immediately suggests protonated acetone $H^+(CH_3)_2CO$ which is rather stable due to the large proton affinity of acetone (193.8 kcal mol⁻¹)⁸. Unfortunately, the bond energy of $H^+(CH_3)_2CO \cdot H_2O$ is not known and, therefore, the second argument cannot be checked. However, bond

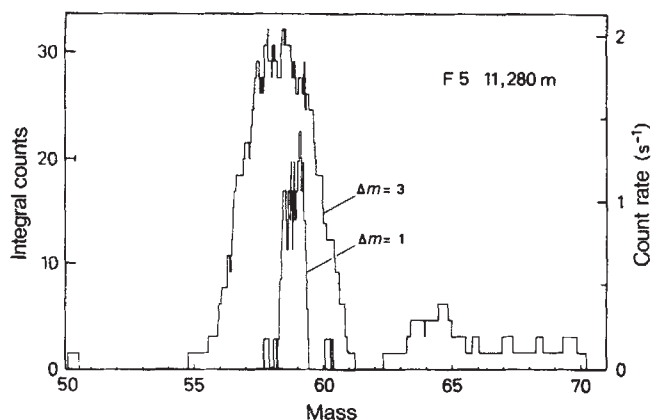


Fig. 2 Portions of positive ion mass spectra with different peak widths at half-peak amplitude, Δm , of 3 and 1 AMU obtained during flight F5 and 11,280 m altitude. Total integration times, 100 and 410 s.

energies for hydrates of large core ions similar to $\text{CH}_3\text{CH}_2\text{CO}^+$ have been measured in the laboratory⁸ and were found to be around 17 kcal mol^{-1} which is, in fact, the same bond energy as $\text{H}_3\text{O}^+(\text{H}_2\text{O})_3$ (ref. 8).

High-pressure flow reactor studies, recently carried out at our laboratory, have revealed that at an ambient pressure of 200 mbar (corresponding to an altitude of around 11,000 m), $\text{H}^+(\text{CH}_3)_2\text{CO}(\text{H}_2\text{O})_n$ clusters are severely dissociated with $\text{H}^+(\text{CH}_3)_2\text{CO}$ being the most prominent fragment ion. It was also found that the fragmentation behaviour was similar to that of $\text{NH}_4^+(\text{H}_2\text{O})_n$ clusters, which are less strongly bonded than $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ clusters (the bond energy for $\text{NH}_4^+\text{H}_2\text{O}$ is only $17.2 \text{ kcal mol}^{-1}$)⁸.

Thus, protonated acetone seems the most likely identification of peak 59. The strongest argument, however, comes from CIMS measurements⁹, also made on flight F5, which reveal the presence of substantial amounts of acetone vapour at 11,280 m (acetone volume mixing ratio, 100 parts per 10^{12} , p.p.t.). Considering the latter acetone abundance one would expect acetone containing clusters to be the most dominant ions, which was in fact the case in F5. From a comparison of the CIMS- and the present ambient ion measurements we can conclude that the $\text{H}^+(\text{CH}_3)_2\text{CO}$ ions are not due to acetone contamination⁹.

If peak 59 was protonated acetone, it seems likely that observed ion families, other than $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$, are also represented primarily by their naked cores ($n=0$). This is to be expected since such ions, including NH_4^+ , usually have much smaller hydration energies compared with H_3O^+ . Thus 55 ± 1 is probably $\text{H}_3\text{O}^+(\text{H}_2\text{O})_2$ and not $\text{NH}_4^+(\text{H}_2\text{O})_2$.

Of the other observed ions (Fig. 1), 33 ± 1 may be identified as protonated methanol or formaldehyde with 50 ± 1 possibly being its hydrate. The ion 28 ± 1 may be attributed to HCO^+ which could be a fragment of protonated acetone⁹. Tentative identifications for the remaining minor peaks are included in Table 1 and will not be further discussed here. Positive ions with masses 29, 31 and 33 are also present in the CIMS spectra of F5 (ref. 9).

From the above, it seems that much more acetone vapour must have been present at 11,300 m in F5 compared with F4. Considering a typical rate coefficient of $2 \times 10^{-9} \text{ cm}^3 \text{ s}^{-1}$ for reaction of acetone with $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$ clusters, and an ion-ion recombination lifetime of 130 s acetone abundances of about 1 p.p.t.v. (F4) and ≥ 20 p.p.t.v. (F5) have been estimated^{1,2}. However, if the ions 37 ± 1 , 55 ± 1 and 73 ± 1 (F4) were not $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$, but $\text{NH}_4^+(\text{H}_2\text{O})_n$ clusters, the above steady-state treatment would not apply.

Acetone vapour measurements at ground level (typically 500 p.p.t.v.) have recently been reported by Penkett¹⁰ using a GC-MS technique. Destruction of acetone (S. Penkett, personal communication) by OH-reaction (lifetime $\sim 1,300 \text{ h}$) and photolysis (lifetime of $\sim 500 \text{ h}$), although being sufficiently slow, allow

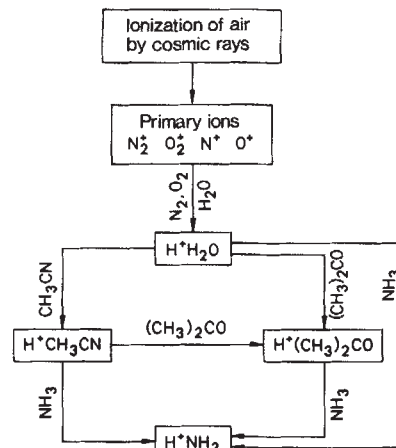


Fig. 3 Simplified reaction scheme for major positive core ions in the upper troposphere and lower stratosphere.

efficient vertical mixing of acetone. However, heterogeneous removal may also operate in the troposphere and, unfortunately, very little is known about this potential acetone sink. Sources of acetone seem to be photo-oxidation of C_3H_8 or higher aliphatic hydrocarbons^{10,11}, preferably through OH-reaction.

Finally, we consider an implication of the present data for lower stratospheric ammonia abundances. If peak 59 is, in fact, protonated acetone, the ammonia abundances at 11,280 m must have been very low for both F4 and F5. This is because $\text{H}^+(\text{CH}_3)_2\text{CO}(\text{H}_2\text{O})_n$ clusters should react efficiently with ammonia. Using a steady-state treatment analogous to the one discussed above, the upper limits to the ammonia abundance at 11,280 m are estimated to be ~ 1 p.p.t.v. (F4) and ~ 0.05 p.p.t.v. (F5).

From the preceding discussion, a simplified picture of the ambient positive-ion chemistry emerges (Fig. 3). Primary ions formed by cosmic ray ionization of air are quickly converted to $\text{H}^+\text{H}_2\text{O}$ 'core' ions which are strongly hydrated ($\sim 6\text{--}7 \text{ H}_2\text{O}$ ligands). Reactions of $\text{H}^+\text{H}_2\text{O}$ 'cores' with the large proton-affinity trace gases acetonitrile and acetone which are almost equal around the tropopause⁹, lead to corresponding $\text{H}^+\text{CH}_3\text{CN}$ and $\text{H}^+(\text{CH}_3)_2\text{CO}$ 'cores'. Incorporation of these molecules into the clusters may also lead to mixed clusters. At lower heights, ammonia vapour^{12,13} becomes abundant enough to convert most of the $\text{H}^+(\text{CH}_3)_2\text{CO}$ 'cores' to H^+NH_3 . Above the tropopause, acetone seems to decrease more rapidly than acetonitrile and ultimately has such a low abundance in the middle stratosphere that $\text{H}^+\text{CH}_3\text{CN}$ conversion becomes inefficient, as suggested by the observed⁴ dominance of $\text{H}^+\text{CH}_3\text{CN}$ cores around 25-km altitude. At even higher altitudes (upper stratosphere), the acetonitrile abundance becomes low enough to prevent efficient transformation of $\text{H}^+\text{H}_2\text{O}$ cores. It is not known at which heights the transitions from dominant H^+NH_3 to $\text{H}^+(\text{CH}_3)_2\text{CO}$ and from $\text{H}^+(\text{CH}_3)_2\text{CO}$ to $\text{H}^+\text{CH}_3\text{CN}$ cores occur.

Future ion composition measurements should be extended downwards in altitude to search for clusters containing H^+NH_3 and other cores with large proton-affinity molecules. Thus *in situ* ion composition measurements may also become a powerful tool for tropospheric trace-gas detection.

We thank E. E. Ferguson, P. Crutzen, D. Ehhalt and D. Griffith for helpful discussions, the technical staff of the MPI-K, particularly K. Oberfrank and the Flugbereitschaft of the DFVLR for assistance. This project was funded by the Bundesministerium für Forschung und Technologie through GSF.

Received 8 May; accepted 9 July 1984.

- Heitmann, H. & Arnold, F. *Nature* **306**, 747–751 (1983).
- Arnold, F., Heitmann, H. & Oberfrank, K. *Planet. Space Sci.* (in the press).
- Arnold, F. & Henschen, G. *Nature* **275**, 521–522 (1978).
- Henschen, G. & Arnold, F. *Geophys. Res. Lett.* **8**, 999 (1981).
- Arijs, E., Ingels, J. & Nevejans, D. *Nature* **271**, 642–644 (1978).
- Ferguson, E. & Arnold, F. *Accts. chem. Res.* **14**, 327 (1981).

7. Arnold, F. in *Atmospheric Chemistry*, (ed. Goldberg, E. D.) (Springer, Heidelberg, 1982).
8. Kebarle, P. A. *Rev. Phys. Chem.* **44S**, (1977).
9. Arnold, F., Oberfrank, K. & Hauck, G. *Planet. Space Sci.* (submitted).
10. Penkett, S. in *Atmospheric Chemistry* (ed. Goldberg, E. D.) (Springer, Heidelberg, 1982).
11. Niki, H. in *Atmospheric Chemistry*, (ed. Goldberg, E. D.) (Springer, Heidelberg, 1982).
12. Georgii, H. W. & Müller, W. J. *Tellus* **16**, 1–2 (1974).
13. Georgii, H. W. & Lenhard, U. *Pageophysics* **116**, (1978).

Microdomains observed at the ferroelectric/paraelectric phase transition of barium titanate

L. A. Bursill & Peng Ju Lin

School of Physics, University of Melbourne, Parkville, 3052, Victoria, Australia

The physical properties of barium titanate have been thoroughly studied because of fundamental interest in its singular ferroelectric properties and the wide variety of applications¹. Inelastic neutron and X-ray diffraction studies suggest that the atomic mechanism for the phase transition from the paraelectric (cubic) to ferroelectric (tetragonal) phases lies somewhere between the extremes of 'displacive' and 'order-disorder' models^{1,2}. Analysis of the 'central-peak' phenomena^{2,3} has led to microdomain models, while molecular dynamic computer studies for ferrodistortive and antiferrodistortive phase transitions predict travelling clusters of locally-ordered regions, with slowing-down and increasing cluster size close to the critical temperature⁴. Direct evidence for the occurrence of microdomains has been obtained by *in situ* electron microscopy of the transition.

Slices (200 μm) of a crystal of BaTiO_3 were mechanically thinned parallel to a pseudocubic face. These were chemically thinned using orthophosphoric acid and sandwiched between two 75-mesh grids. Specimens were observed in a Gatan double-tilting heating-cooling goniometer ($-180 < T < 150^\circ\text{C}$) in a JEOL-100CX electron microscope. Images were recorded as a function of temperature using standard photographic techniques as well as a television camera with videotape recorder.

The room-temperature observations closely followed the results of Tanaka and Honjo⁵. Thus 90° domains of a-c and a-a type occurred frequently (see ref. 6, Fig. 2). On heating above $120\text{--}130^\circ\text{C}$, a faint contrast appeared between existing domain walls (Fig. 1). Approximately equal areas of light and dark contrast occurred, with periodicity $\sim 425 \pm 25 \text{ \AA}$. Dots of contrast and wavy lines occurred with ill-defined orientation and connectivity. This 'microdomain' contrast could be enhanced (Fig. 1) by tilting off the $[001]$ zone axis by a few degrees and using the $hh0$ systematic row diffraction condition (L.A.B. and P.J.L., in preparation). For a wedge-shaped crystal (Fig. 2) the microdomain periodicity varied systematically with crystal thickness, ranging from $\sim 250 \text{ \AA}$ in the thinnest areas, to $\sim 1,200 \text{ \AA}$ for a crystal thickness of $\sim 1,000 \text{ \AA}$. Displacing or focusing the electron beam did not change the gross features of the pattern and no sideways movements of the microdomains could be induced. This behaviour is quite distinct from that of a-c or a-a ferroelectric domain walls, which could be induced to move sideways quite readily by moving or focusing the electron beam. Closer examination of videotapes showed fluctuations within microdomain contrasts (see arrowed positions in Fig. 3). Thus, for specimens thermally equilibrated at temperatures in the range $130\text{--}140^\circ\text{C}$, small patches of darker contrast ($\sim 40\text{--}60 \text{ \AA}$ diameter) could be seen to transfer across the domain walls between adjacent dark contrast spots or stripes. The overall periodicity and texture of the microdomains were maintained, but they clearly possessed a dynamic character on a finer ($20\text{--}40 \text{ \AA}$) scale. Such movements, detectable at 25 frames s^{-1} , probably contribute to the diffuse appearance of the microdomain wall contrast. Such fluctuations were not simply due to small

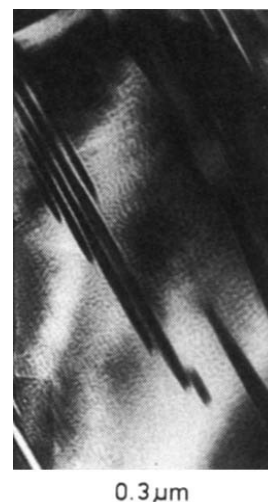


Fig. 1 Microdomain contrast within ferroelectric domains in a BaTiO_3 crystal observed at $120\text{--}130^\circ\text{C}$ (pseudocubic projection axis). Note equal areas of light and dark contrast with periodicity $\sim 425 \text{ \AA}$.

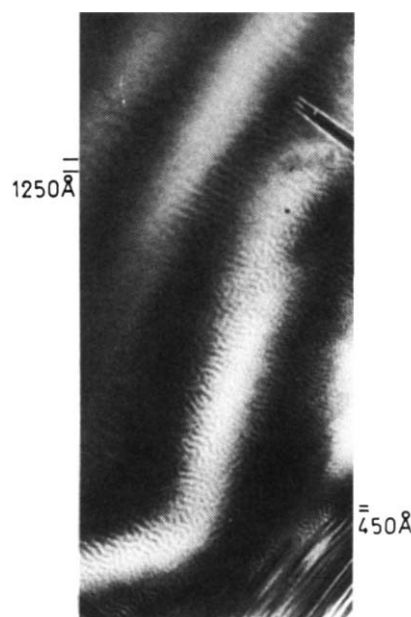


Fig. 2 Thickness dependence of microdomain periodicity in wedge-shaped crystal of BaTiO_3 (pseudocubic projection).

tilts during observation, as judged from the stability of bend extinction contours.

Attempts were made to establish systematic relationships between microdomain size and/or fluctuation frequency and temperature. These were not conclusive. On cooling the specimens from 140°C to room temperature, the microdomains sometimes disappeared at $\sim 100\text{--}120^\circ\text{C}$ but sometimes remained stable to $\sim 60\text{--}80^\circ\text{C}$. They did not occur at room temperature. The stability of microdomains on cooling appeared to be related to the topology of the specimen, in particular to elastic bending, and to the cooling rate. There was no systematic variation of domain size on cooling, and the microdomains probably disappear suddenly rather than gradually. The onset temperature was more sharply defined, being 127°C for three different experiments, indicating reproducible behaviour. In other experiments the onset temperature varied between 130 and 150°C . Thus exceptionally flat and extensive wedge-shaped thin edges, which were initially free of a-c or a-a domain walls, were strongly

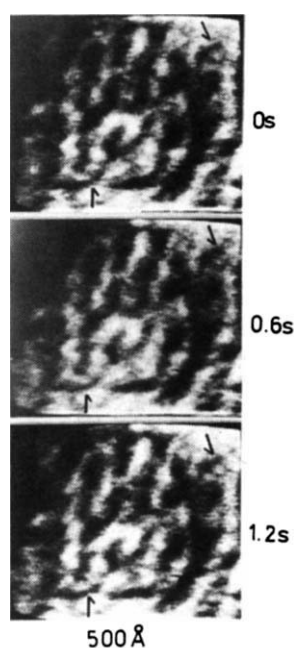


Fig. 3 Time sequence from videotape recording showing fluctuations in microdomains at 140 °C (25 frames per s). Note transfer of small (50-Å diameter) areas of darker contrast between adjacent microdomains, whereas the overall domain periodicity and topology is essentially maintained.

resistant to microdomain nucleation, even up to the maximum available temperature (150 °C).

Microdomain nucleation was clearly demonstrated for such a wedge-shaped crystal which contained only a single a-c domain wall. This wall remained quite stable until 140 °C and then suddenly diverged to create a slab containing microdomains bounded on either side by fault fringe contrast (Fig. 4a). Slight changes in the incident beam illumination angle were then enough to cause this slab to expand or contract, following movement of the incident beam. The microdomain texture often showed topology apparently determined by that of existing a-c or a-a domains. Thus for Fig. 4a the microdomains form essentially a one-dimensional array, parallel to the existing (011) wall, whereas in Fig. 4b some elements of two-dimensional short-range order appear to be imposed by orthogonal (011) and (01 $\bar{1}$) a-c domain walls. The fine scale of the latter sometimes appears as a precursor to microdomains formation. In other areas (see Figs 1 and 2), the microdomains do not show preferred orientations.

Dark-field observations of the microdomains will be discussed elsewhere (unpublished work), as will other results for domain wall kinetics, recorded for $-180\text{ °C} \leq T \leq 150\text{ °C}$. These also concern the tetragonal/orthorhombic (5 °C) and orthorhombic/trigonal phase transitions. Distinctive behaviour often occurred at thin edges and cracks.

The observed mechanism of microdomain formation, by a-c or a-a wall divergence, and the fine-scale fluctuations involving transfer of one-ferroelastic orientation between adjacent microdomains should add more reality to molecular dynamical⁴ and other theoretical approaches to the understanding of structural phase transitions².

The structure of the microdomains and hence contrast mechanisms have not been determined. Five important new observations need to be explained: (1) apparently equal volumes of bright and dark contrasts appeared in every case recorded; (2) the microdomains did not readily interact with the electron-beam, whereas a-c and a-a domain walls in the tetragonal phase move readily on focusing or translating the incident beam; (3) the microdomain contrast is relatively very weak compared with

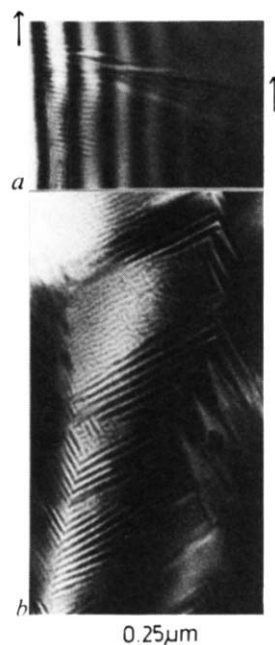


Fig. 4 a, Divergence of a-c domain wall at 140 °C, creating a slab containing microdomains bounded by inclined boundaries. Note one-dimensional nature of microdomains, lying parallel to existing a-c wall. b, Elements of two-dimensional short-range ordering of microdomains, apparently imposed by topology of coexisting a-c ferroelectric domain walls.

normal 90° domain wall contrast; (4) the microdomain walls are relatively diffuse and subject to fine-scale fluctuations; and (5) microdomains may coexist with normal a-c or a-a domain walls, often showing topology apparently related to the latter.

The following tentative interpretation of the structural mechanism for the ferroelectric/paraelectric phase transition is based on these observations. We propose that the bright and dark microdomains contain a structure intermediate between that of the well-known¹ tetragonal and cubic forms of BaTiO₃. This transition structure could simply be a tetragonal distortion of the cubic phase. This would allow domain formation and give reasonable changes in background contrast across the microdomain walls (L.A.B. and P.J.L., in preparation). It should also be paraelectric, so that there is no electrical interaction between the incident beam and the microdomain walls. The transitional phase could also be ferroelastic, with noncentrosymmetric displacements of the oxygen atoms with respect to the cubic phase, but retaining zero electric dipole moment. Such a structural mechanism is represented in Fig. 5. Clearly, the tetragonal distortion may take different orientations with respect to the cubic (or tetragonal) axes, but to minimize elastic strain overall, equal volumes of symmetrically equivalent domains should occur. The coexistence of a-c or a-a ferroelectric domains in the tetragonal phase, together with microdomains of the transitional phase (Fig. 4b), occurs over a range of temperatures, and a strong influence of the former on the topology of the latter is to be expected, as the structural difference involves small atomic displacements within virtually identical tetragonal unit cells (Fig. 5).

Our observations (Fig. 3) give some insights into the question of microdomain dynamics. The observation of a sharp onset temperature for microdomain formation (Fig. 3), rather than a gradual coarsening with increasing temperature, suggests that some modification of the models used for molecular dynamical calculations is required.

The present work has obvious limitations. First, the restriction to $T \leq 150\text{ °C}$ meant that it was not possible to study the stability of microdomains within the cubic phase. The molecular dynamic computations may then be more appropriate, with behaviour

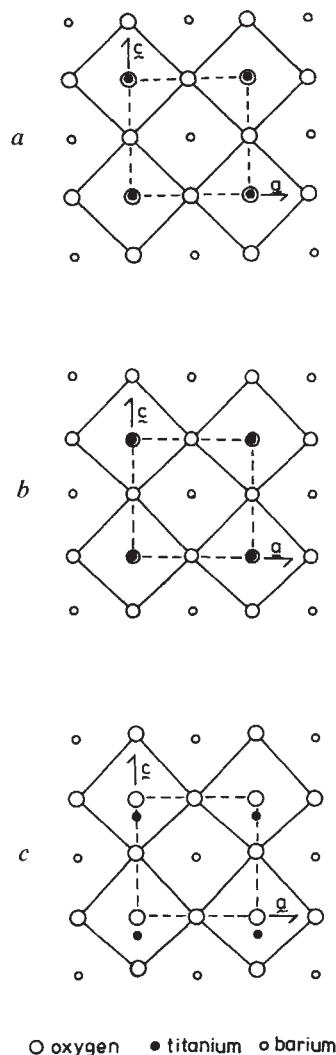


Fig. 5 Structural relationships between cubic (a), transitional (b) and tetragonal (c) phases of BaTiO_3 .

quite different from that within the tetragonal phase. Second, the observed thickness dependence of microdomain periodicity (Fig. 2) indicates that at least some of the observed behaviour may not be characteristic of the bulk. Further studies using higher voltage instruments are required.

Clearly, transmission electron microscopy may be used to study some aspects of critical fluctuation phenomena associated with structural phase transitions. The present *in situ* studies have provided insights into the ferroelectric/paraelectric transition in barium titanate and the techniques used should have much wider applicability, even given that there may be some limitations, especially with regard to the applicability of the results for bulk crystals. However, many future device applications, and hence fundamental interests, are likely to involve thin films.

This work was supported by the Australian Research Grants Committee and the University of Melbourne. P.J.L. thanks the Education Department of China for permission to visit Melbourne. We thank Dr A. E. C. Spargo and P. Hanan (University of Melbourne) for assistance.

Received 12 June; accepted 14 August 1984.

1. Lines, M. E. & Glass, A. M. *Principles and Applications of Ferroelectrics and Related Materials* Ch. 8 (Clarendon, Oxford, 1977).
2. Cowley, R. A. & Bruce, A. D. *Adv. Phys.* **29**, 1–321 (1980).
3. Müller, K.-A. *Non-Linear Phenomena at Phase Transitions and Instabilities* (ed. Riste, T.) 1–34 (Plenum, New York, 1981).
4. Schneider, T. & Stoll, E. *Phys. Rev.* **B13**, 1216–1237 (1976); **B17**, 1302–1322 (1978).
5. Tanaka, M. & Honjo, G. *J. phys. Soc. Jap.* **19**, 954–970 (1964).

Invariant and changing transitional field configurations in a sequence of geomagnetic reversals

Jean-Pierre Valet & Carlo Laj

Centre des Faibles Radioactivités, Domaine du CNRS,
Avenue de la Terrasse, 91190, Gif-sur-Yvette, France

In the past few years, the increasing number of detailed records of geomagnetic transitions has provided new and important information about the behaviour of the Earth's magnetic field during a reversal. Surprisingly, however, no attempt has been made to obtain information about the evolution of the transitional field characteristics over a long period of time. We report here the results of a palaeomagnetic study of four different geomagnetic reversals, three of which are successive, recorded at different stratigraphical levels of two Upper Tortonian marine clay sections in western Crete (Greece). This sequence represents ~ 1.5 Myr of geomagnetic field history. The results obtained suggest that the process of reversal from either starting polarity may remain invariant for a long period of time, then change rather abruptly between successive reversals.

The two sections investigated are situated 5 km apart in the Kastelli basin in western Crete, near the villages of Potamida and Skouloudhiana. A careful correlation between these two sections, among others, has been established by a previous extensive bio- and magnetostratigraphical study of late Miocene sediments in Crete¹. The location of the two sections is shown schematically in Fig. 1, together with the corresponding correlated polarity sequences. Correlation with the geomagnetic time scale shows that the two sections combined cover the time interval between 7.4 and 5.3 Myr, and that the four reversals studied here are spread over a time interval of ~ 1.5 Myr (ref. 1).

Some of the results obtained from these two sections have already been reported^{2–4}. We add here new results for the normal-reversed (N–R) KS 05 transition. Moreover, extensive resampling has allowed us to obtain further transitional palaeomagnetic directions, so that all the palaeomagnetic records may be classified as category A, as defined by Fuller *et al.*⁵.

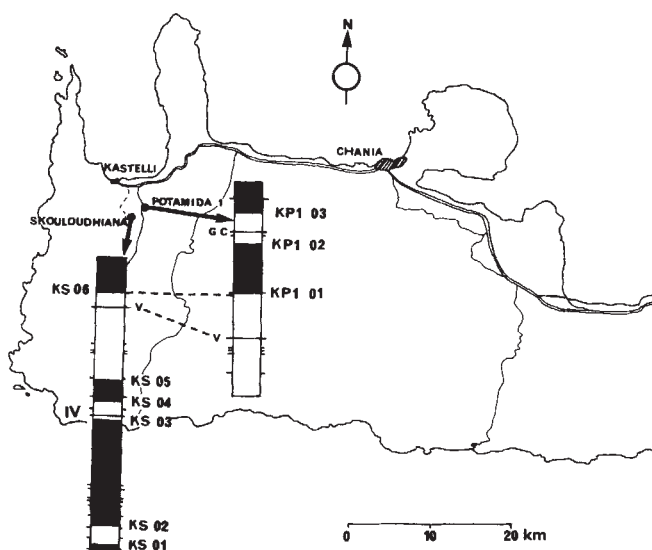


Fig. 1 Map showing the location of Potamida and Skouloudhiana sections together with the biomagnetostratigraphical correlation of the corresponding polarity sequences adapted from ref. 1. GC, First occurrence datum (FOD) of *G. conomiozea* group; V, FOD of *G. menardii* form 5; IV, last occurrence datum of *G. menardii* form 4.

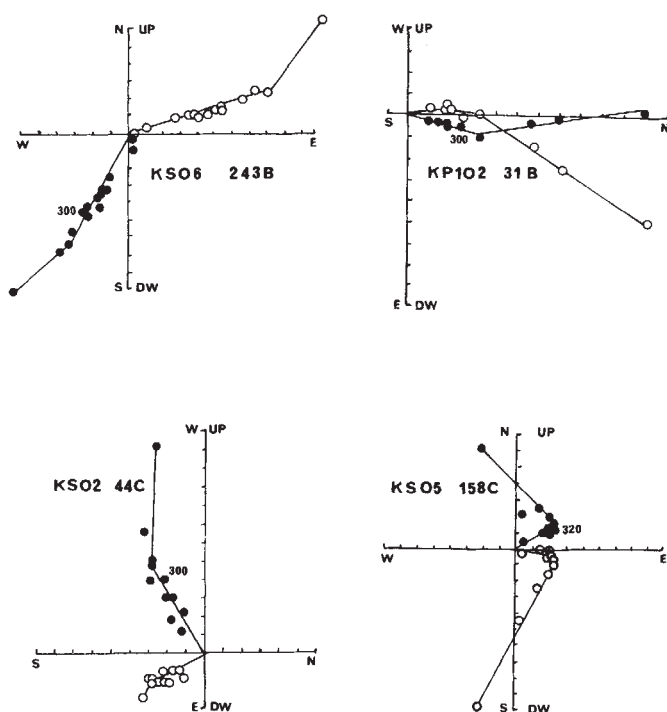


Fig. 2 Typical thermal demagnetization diagrams of samples from KS 02, 05, 06 and KP1 02 transition zones. ●, Declination data; ○, inclination data.

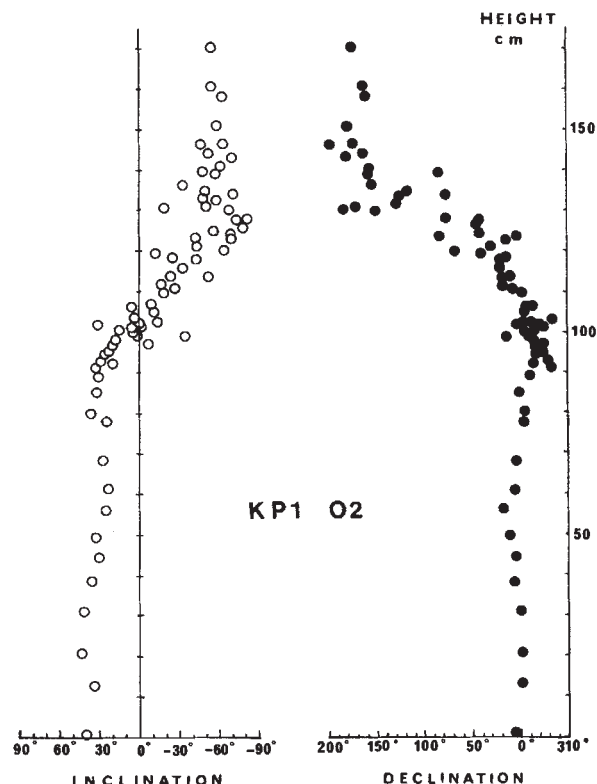


Fig. 3 Detailed declination/inclination record relative to the (N–R) KP1 02 transition in the Potamida 1 section.

The sampling was done using standard palaeomagnetic techniques, with either water or compressed air as a coolant. Care was taken to determine the exact relative stratigraphical position of the cores in each transitional zone. Repetitive measurements yielded identical results within 0.5 cm. The exact stratigraphical height of the different samples obtained from each core was calculated using a computer program which took into account the usual orientation parameters of the core, its stratigraphical position and even the width of the diamond saw used to cut the cores. The magnetic properties of the clays have been carefully investigated to check that there are no changes in the magnetic or sedimentary mineralogy over the different transitional zones.

Isothermal remanent magnetization (IRM) measurements show that, in all cases, magnetite is the main magnetic mineral, with a very small amount of a mineral of higher coercivity, presumably haematite, also generally present.

The anisotropy of the magnetic susceptibility has also been measured at different horizons of the two sections. In both cases, the magnetic fabric is described by an oblate ellipsoid with a very moderate (1.05) degree of anisotropy. In Skouloudhiana, the mean $K_{\min}$ axis is perfectly perpendicular to the bedding plane (inclination $I = 88.5^\circ$) while the $K_{\max}$ axes of the individual samples show no preferential orientation in this plane. In Potamida 1, the well grouped $K_{\min}$ axes dip slightly in a northeasterly direction and there is a preferential north–south trend of the $K_{\max}$ axes. The magnetic lineation $(3(K_{\max} - K_{\min}) / (K_{\max} + K_{\min} + K_{\text{int}}))$ is, however, very small (0.014), so that in both sections the results indicate a planar undisturbed sedimentary fabric.

Stepwise thermal demagnetization of all samples has been used and the magnetization was measured with either a Digico spinner magnetometer or a three-axis LETI cryogenic magnetometer. For the most weakly magnetized samples, and in particular for all the samples of the KS 05 transition, the double precision procedure of Vandenberg⁶, involving four independent measurements for each sample, was used with the

cryogenic magnetometer. In spite of this procedure, some of the demagnetization diagrams can hardly be judged to be linearly decreasing. These diagrams which represent ~30% of the total in KS 05, were not taken into account in the final results. No such problems were encountered for the other transitions.

Typical demagnetization diagrams of samples from four different reversals are shown in Fig. 2; these are plotted relative to transitional samples so that they correspond to the lowest natural remanent magnetization (NRM) and most delicate part of each transition. It can be seen that, apart from a secondary, probably viscous, component removed at low temperature ($\leq 250^\circ\text{C}$), only a single stable component of magnetization is present.

The magnetic susceptibility of samples from each transition was monitored during progressive demagnetization using a susceptibility bridge. No changes of more than a factor of three were found between room temperature and 500°C , indicating that no major mineralogical transformations had affected the magnetic materials.

Palaeomagnetic directions were obtained by hand fitting a straight line through the last points of the demagnetization plots. As an example, the complete declination–inclination record for transition KP1 02 is shown in Fig. 3.

In all the sediments, the mean inclination of the samples situated well outside the transition zones is about $8\text{--}10^\circ$ shallower than that expected at the site latitude on the basis of a geocentric axial dipole. This is a general property, not only of Potamida and Skouloudhiana^{1,2}, but also of all sediments of the same nature throughout the Hellenic sedimentary arc⁷. A correction of this error using a formula such as $\tan I' = f \tan I$ (ref. 8) does not, however, significantly change estimates of the virtual geomagnetic pole (VGP) paths² and we have therefore worked with uncorrected measurements.

From the directions obtained from the demagnetization diagrams, the successive VGP positions have been calculated for the five transitions. The results are shown in Fig. 4 together

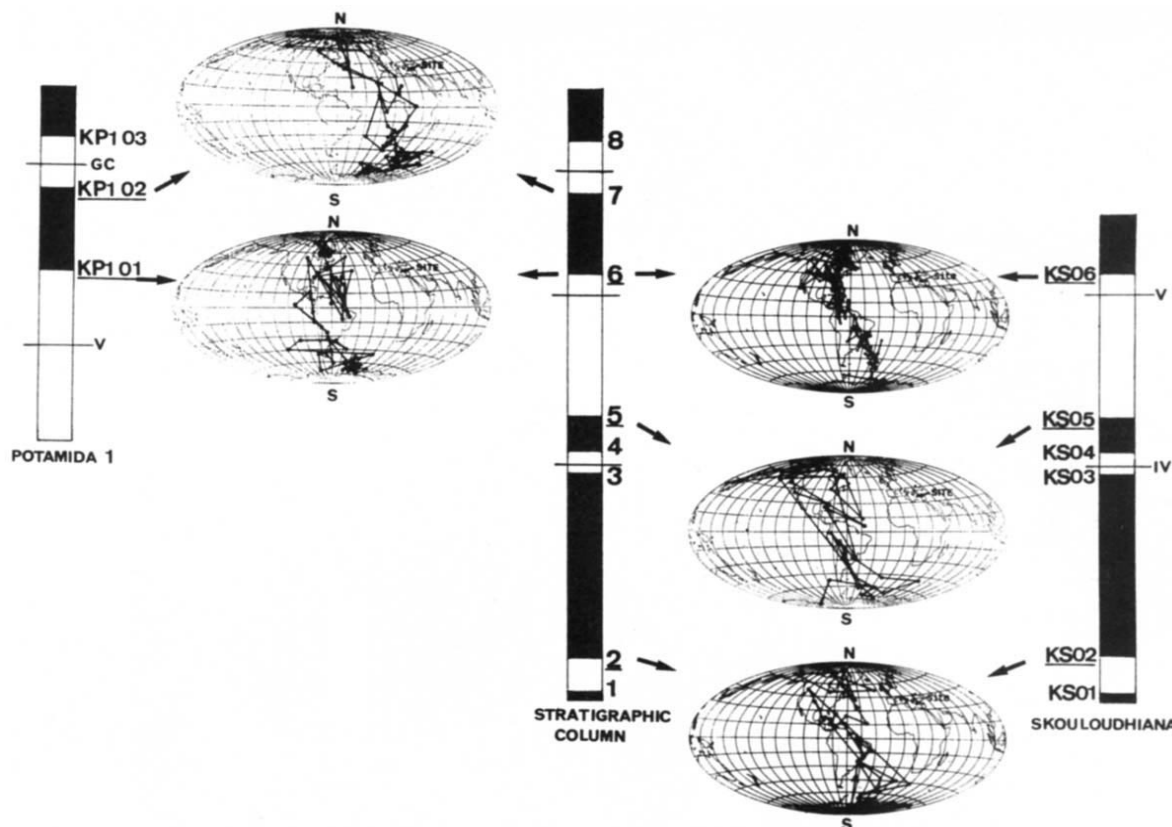


Fig. 4 VGP paths of the five polarity transitions (north VGPs for the R–N and south VGPs for the N–R transitions). The composite total stratigraphical column in the centre gives the stratigraphical position of the transitions. The VGP paths of the lower three transitions are identical within noise limits, while the upper one differs by $\sim 60^\circ$ long.

with the stratigraphical position of the reversals. To compare directly the VGP paths corresponding to transitions of opposite sense, we have plotted in Fig. 4 the north VGPs associated with R–N transitions and the south VGPs associated with the N–R transitions, as suggested by Hoffman⁹.

It can be seen from Fig. 4 that, within noise limits, the VGP paths associated with the transitions KS 06 and KP1 01, which represent the record of the same transition in the two sections, are identical. This, of course, is a condition that must be fulfilled if transitional VGP paths from the two sections are to be compared; the fact that it is suggests that differences in the VGP paths of other transitions are probably a result of changes in the reversal mechanism and not of distortion in the sampled sections.

The main result (Fig. 4) is that the north–VGP paths associated with the two (R–N) transitions, and the south–VGP path for the (N–R) transition in Skouloudhiana, are identical. In spite of significant scatter, which is a consequence of the low magnetization of the sediments for the two lower transition zones, even an east to west migration of the intermediate VGPs can be recognized in these plots as already noted by Hoffman⁹. The directional field variation at the site during this (N–R) transition is thus identical to that associated with the R–Ns except for a change in sign. (This situation is exactly opposite to a standing field situation which requires identical transitional field along with sign, but is consistent with a flooding model.)

On the other hand, the south–VGP path relative to the upper (N–R) transition in Potamida 1 is significantly different, being about 60° of longitude away from the other paths.

The marine sediments used in this study have an average sedimentation rate of $\sim 4\text{--}5\text{ cm kyr}^{-1}$ (ref. 1). Therefore, they may not record changing field directions as faithfully as rapidly deposited lake sediments or, indeed, volcanic rocks. These sediments have, however, proved to be a faithful record of the direction of the steady magnetic field to within a few degrees^{1,7}. It is, thus, very difficult to imagine that poor remanence acquisi-

tion processes for a rapidly changing field, such as long integration times for the locking-in of the remanence, would modify the general features of the reversal records. In particular, we cannot think of any process which would result in antipodal reversal paths for the different transition senses. This feature, in contrast, can be taken as proof of the reliability of the palaeomagnetic records.

These results, therefore, suggest that the process of reversal from either starting polarity remained unchanged during the time interval of 1.2–1.3 Myr, corresponding to the first three reversals¹. A change then occurred in the period between the two upper reversals. This change must have occurred rather abruptly, because these two reversals are separated by $<250,000\text{ yr}$ (ref. 1).

Similarities, or even coincidence, of VGP paths of different transitions, recorded in different, but geographically close, sites, have already been reported^{10–12}. Hoffman has even suggested, on the basis of four Cenozoic transitions in the western US that because of particular conditions in the core, only a very limited number of possible reversal configurations might be liable to occur with finite probability at any epoch, so that only particular ones will be frequently observed in time¹³. The results reported here are, however, the first to suggest that the mechanism of reversal is invariant over a long period of time, confirming previously reported results³.

Moreover, these results from the upper Tortonian epoch cannot be reconciled with the standing field model. It is thus interesting that a set of two successive palaeomagnetic reversals obtained by Bogue and Coe from slightly younger (5.6–4.5 Myr) lava flows in Kauai yield, in contrast, very similar paths¹⁴. These authors have noted that unless the standing field or the flooding process can, at least occasionally, change from reversal to reversal, neither model can explain all the available transition data. The present results, together with those from Kauai, suggest that such changes in the reversal process might occur not only from reversal to reversal but also between different epochs,

so that a particular process might be more effective during a particular period of time.

In this respect, the results from the (R–N) KPI-03 transition would be quite interesting. Indeed its age is 5.6 Myr (ref. 1), so that it could eventually provide a link between the results from Crete and those from Kauai. Unfortunately the sediments in the KPI 03 transition zone are weakly magnetized and furthermore a ferruginous level is present in this zone¹. Thus, such measurements would be quite difficult to achieve.

We thank Cor Langereis for invaluable field assistance, H. Zijderfeld for many helpful discussions, and K. Hoffman for detailed comments and suggestions. The Director General of the Institute of Geological and Mining Research (IGME) in Athens provided the necessary permits. C. Kissel, S. Sen and M. Villegier helped with the sampling. This work was supported by the French CNRS ATP Géologie et Géophysique des Océans.

Received 24 April; accepted 20 July 1984.

- Langereis, C. G., Zachariasse, W. J. & Zijderfeld, J. D. A. *Mar. Micropaleontol.* **8**, 261–281 (1983–84).
- Valet, J. P., Laj, C. *Earth planet. Sci. Lett.* **54**, 53–63 (1981).
- Valet, J. P., Laj, C. & Langereis, C. G. *Nature* **304**, 330–332 (1983).
- Valet, J. P. & Laj, C. *Nature* **309**, 90–91 (1984).
- Fuller, M., Williams, I. & Hoffman, K. A. *Rev. Geophys. Space. Phys.* **2**, 179–203 (1979).
- Vandenbergh, J. *Tectonophysics* **98**, 29–41 (1983).
- Laj, C., Jamet, M., Sorel, D. & Valente, J. P. *Tectonophysics* **86**, 45–67 (1982).
- King, R. F. *Mon. Not. R. astr. Soc. Geophys. Suppl.* **7**, 115 (1955).
- Hoffman, K. A. *Nature* **309**, 90 (1984).
- Mankinen, E. A., Donnelly-Nolan, J. M., Gromme, C. S. & Hearn, B. C. Jr. *U.S. geol. Surv. Prof. Pap.* **1141**, 67–82 (1981).
- Liddicoat, J. C. *Phil. Trans. R. Soc. A* **306**, 121–128 (1982).
- Dodson, R. et al. *Geophys. J. R. astr. Soc.* **373**–412 (1978).
- Hoffman, K. A. *Earth planet. Sci. Lett.* **44**, 7–17 (1979).
- Bogue, S. W. & Coe, R. S. *Nature* **295**, 399–401 (1982).

Density of the ocean crust

R. L. Carlson & G. S. Raskin

Department of Geophysics and Geodynamics Research Program,
Texas A&M University, College Station, Texas 77843, USA

The density structure and average density of the oceanic crust have implications for various geological and geophysical problems, including interpretations of gravity data, the variation of lithospheric buoyancy in relation to age, and role and fate of the crust in subduction. But no systematic evaluation of oceanic crustal density has been made and estimates range from 2.85 to 3.0 Mg m⁻³ (refs 1–3). We have made an evaluation based on seismic refraction data in combination with drilling results, laboratory studies of seismic properties of oceanic and ophiolitic rocks, and ophiolite lithostratigraphy. Our preferred value for the mean density of the oceanic crust is 2.89 ± 0.04 Mg m⁻³.

Seismic refraction data provide the most important constraints on the structure and composition of the oceanic crust. (Seismic models and their interpretation in terms of composition are given in refs 4, 5.) Over the past decade or so, simple three-layer velocity-structure models have given way to more detailed multi-layer models, and synthetic seismogram modelling has led to more sophisticated 'gradient' models.

Synthetic seismogram models indicate that the upper 2 km of igneous basement (layer 2) is characterized by strong vertical velocity gradients, with P-wave velocities ranging from <4 km s⁻¹ at the top to >6 km s⁻¹ at the bottom of the interval. Layer 2 may, nevertheless, be subdivided into layers (2a, 2b, 2c), though these regions may not be homogeneous or separated by distinct discontinuities^{5,6}. The rapid increase of velocity with depth in layer 2 may be attributed to several factors: (1) velocities in the upper 200–400 m of the section are significantly reduced by large-scale porosity (produced, for example, by collapsed pillows, and lava tubes)^{7–9}; (2) there is a systematic decrease of grain-boundary porosity with depth^{10–12}; and (3) metamorphic

grade increases with depth; greenschist facies mineral assemblages (chlorite, epidote, actinolite, albite) are associated with layer 2c in hole 504B (ref. 9). In contrast to layer 2, velocity gradients in oceanic layer 3 are low (0.0–0.1 S⁻¹). The velocity at the top of the unit is typically 6.7–6.8 km s⁻¹, but the internal structure of layer 3 is apparently variable. In some cases, the gentle gradient extends to the Moho; in others the layer may be subdivided into two (or more) units^{4,5,13}. The transition from crustal to mantle velocities generally occurs over an interval of ~1 km.

The oceanic crust can be approximated by a layered sequence, having an average density given by

$$\rho_{\text{avg}} = \frac{1}{T} \sum_{i=1}^n \rho_i t_i; T = \sum_{i=1}^n t_i \quad (1)$$

where ρ_i and t_i are the density and thickness of the i th layer, respectively, n is the total number of layers and T is the total thickness of the sequence. The number of layers and their thicknesses may be taken directly from seismic velocity-structure models; the problem is to find a means of estimating the layer densities. We have approached this problem in three ways.

Some ophiolites, at least, are now believed to be slices of oceanic crust and upper mantle. The velocity structures of the Semail Ophiolite, Oman and the Bay of Islands Complex, Newfoundland, for example, have been reconstructed on the basis of laboratory studies^{14–16}. The principal lithologies present in the 'crustal' sections of ophiolites are basalts, diabases, gabbros and their metamorphic equivalents. A much simplified lithologic sequence, viewed from the top downward, consists of extrusive, generally pillowed, basalts (1–1.5 km thick), sheeted diabase dykes (1–1.5 km), and a high-level isotropic gabbro sequence (up to 1 km), which grades downward into a 3–5-km thick sequence of mafic, layered gabbros (hornblende, pyroxene, and olivine gabbro, troctolite, and so on) in which the ultramafic component typically increases with depth. The mafic gabbros are commonly, but not always, interlayered with ultramafic cumulates in the lower part of the 'crustal' section.

The velocity structures of ophiolites are similar to the velocity structure of the oceanic crust^{14,15,17}. In ophiolites, 'layer 2' is characterized by strong velocity gradients which extend well into the sheeted dykes and reflect the combined effects of a systematic increase in metamorphic grade (zeolite to upper greenschist facies) and a systematic decrease in grain-boundary porosity with depth^{14–17}. Velocities typical of oceanic layer 3 are encountered in the middle to lower part of the dyke sequence, where upper greenschist assemblages give way to those of lower amphibolite grade. The velocity gradient below this level is relatively low, and differences between ophiolites in the amount of modal olivine present in the lower part of the gabbro sequence account for the observed variability of layer 3 velocity structures¹⁶.

In method 1, to estimate crustal density, we assume that layer 2 consists of basalts and dyke rocks in about equal proportions. The upper part of layer 3 is taken to consist of gabbro and metagabbro, with denser, olivine gabbros occurring in the lower half the layer. Neglecting the effect of interlayered ultramafic rocks in the lowermost part of the crustal section, we have used

Table 1 Densities of ophiolite samples

Lithology	No. of samples	Density (Mg m ⁻³)
Metabasalt, basalt and spilite (B)	34	2.73 ± 0.09
Diabase and metadiabase (D)	31	2.85 ± 0.07
Metagabbro (MG)	27	2.90 ± 0.07
Gabbro and olivine gabbro (G + OG)	38	2.95 ± 0.07
Metagabbro, gabbro and olivine gabbro (G + MG + OG)	65	2.93 ± 0.07

Table 2 Summary of average velocity-structure models and density structures

Ref.	Model		Method 1		Method 2	Method 3
	Layer	Thickness (km)	Velocity (km s ⁻¹)	Lithology*	Density (Mg m ⁻³)	Density (Mg m ⁻³)
24	2	1.71 ± 0.75	5.07 ± 0.63	B + D	2.79 ± 0.08	2.75 ± 0.09
	3	4.86 ± 1.42	6.69 ± 0.26	G + MG + OG	2.93 ± 0.07	2.91 ± 0.04
	Mean				2.89 ± 0.06	2.87 ± 0.04
25	2	1.49 ± 0.98	5.19 ± 0.69	B + D	2.79 ± 0.08	2.77 ± 0.10
	3	4.62 ± 1.30	6.81 ± 0.16	G + MG + OG	2.93 ± 0.07	2.93 ± 0.03
	Mean				2.90 ± 0.06	2.89 ± 0.04
4	2	1.39 ± 0.5	5.04 ± 0.69	B + D	2.79 ± 0.08	2.75 ± 0.10
	3	4.97 ± 1.25	6.73 ± 0.19	G + MG + OG	2.93 ± 0.07	2.92 ± 0.04
	Mean				2.90 ± 0.06	2.88 ± 0.04
4 Sonobouy Type I	2	1.6	4.4	B + D	2.79 ± 0.08	2.64 ± 0.01‡
	3a	1.2	6.4	MG	2.90 ± 0.07	2.91 ± 0.01
	3b	4.8	7.1	G + OG	2.95 ± 0.07	2.97 ± 0.03
	Mean				2.91 ± 0.05	2.89 ± 0.02
4 Sonobouy Type II	2	1.6	4.4	B + D	2.79 ± 0.08	2.64 ± 0.01‡
	3a	3.0	6.8	MG	2.90 ± 0.07	2.93 ± 0.03
	3b	2.6	7.5	G + OG	2.95 ± 0.07	3.01 ± 0.01
	Mean				2.89 ± 0.04	2.89 ± 0.01
6 Atlantic	2a	0.3 ± 0.1	3.74 ± 0.50	B	2.73 ± 0.09	2.49 ± 0.14
	2b	1.0 ± 0.1	5.13 ± 0.38	B + D	2.79 ± 0.08	2.76 ± 0.06
	2c	1.0 ± 0.3	6.05 ± 0.22	D	2.85 ± 0.07	2.87 ± 0.03
	3	4.97 ± 1.25†	6.83 ± 0.21	G + MG + OG	2.93 ± 0.07	2.93 ± 0.04
	Mean				2.89 ± 0.05	2.88 ± 0.03
6 Pacific	2a	0.4 ± 0.1	3.47 ± 0.35	B	2.73 ± 0.09	2.41 ± 0.11
	2b	0.8 ± 0.1	5.28 ± 0.39	B + D	2.79 ± 0.08	2.78 ± 0.05
	2c	0.9 ± 0.4	6.12 ± 0.18	D	2.85 ± 0.07	2.88 ± 0.02
	3	4.97 ± 1.25†	6.90 ± 0.17	G + MG + OG	2.93 ± 0.07	2.94 ± 0.03
	Mean				2.89 ± 0.05	2.89 ± 0.03
13	2a	0.38	5.30	B + D	2.79 ± 0.08	2.78 ± 0.01‡
	2b	1.93	6.10	D	2.85 ± 0.07	2.88 ± 0.01
	3a	1.55	6.86	MG	2.90 ± 0.07	2.94 ± 0.03
	3b	3.15	7.06	G + OG	2.95 ± 0.07	2.96 ± 0.03
	Mean				2.90 ± 0.04	2.92 ± 0.01

* Abbreviations as in Table 1.

† Thickness of layer taken from the average structure of Christensen and Salisbury⁴.

‡ Calculated standard deviations do not include the effect of variability in crustal structure.

published water-saturated bulk densities¹⁴⁻¹⁷ to estimate the mean densities of pertinent combinations of rock types given in Table 1.

Assigning densities to velocity structure models on the basis of ophiolite lithostratigraphy neglects the large-scale porosity of layer 2a (refs 7, 8), and is based on a lithologic model which is only loosely tied to the velocity structure. A preferable approach is one in which layer densities can be calculated from a statistically-valid, functional relationship between formation density and seismic velocity.

Assuming that the variation of grain properties is small relative to the difference between the properties of the grains and those of the pore medium (seawater), we may combine the theoretical density and time-average relations for two-component systems, to obtain a velocity/density function

$$\rho = a + b/V_p \quad (2)$$

Fitting this equation to the laboratory densities and P-wave velocities (measured at 100 MPa) of samples recovered by the Deep Sea Drilling Project and from ophiolites¹⁴⁻²¹ (Fig. 1) yields

$$\rho(\text{Mg m}^{-3}) = (3.81 \pm 0.02) - (5.99 \pm 0.11)/V_p(\text{km s}^{-1}) \quad (3)$$

with a coefficient of determination (r^2) of 0.86 and an r.m.s. error of 0.07 Mg m⁻³. Equation (3) describes the velocity/density relation for oceanic rocks very well.

In method 2 layer densities are calculated from seismic velocities according to equation (3). The relation does not account for increasing confining pressure, but the effect of

pressure (above 100 MPa) on density and velocity is relatively small. Equation (3) reflects the effects of grain-boundary porosity and metamorphic grade, but also includes effects of alteration¹⁸⁻¹⁹, which is not representative of oceanic basalts. This method also fails to account for large-scale porosity. Thus method 2 may be applied to deeper levels of the crust where porosity is low, but cannot be applied with confidence to oceanic layer 2.

Seismic velocities in porous media depend on the shapes of the voids as well as the properties of the matrix and pore medium. Because the shapes of large-scale voids in layer 2 are unknown, there are likely to be large uncertainties in the any postulated relationship between velocity and density or porosity⁵. We have, therefore, taken the simplest possible approach of calculating the coefficients of equation (2) from the properties of an average, zero-porosity basalt and the properties of seawater ($\rho = 1.025 \text{ Mg m}^{-3}$; $V_p = 1.5 \text{ km s}^{-1}$). The average bulk-grain densities of oceanic basalts range from 2.9 to 3.0 Mg m⁻³ (refs 7, 10, 12); we estimate the properties of a typical, zero-porosity basalt to be $\rho = 2.03 \pm 0.06 \text{ Mg m}^{-3}$ and $V_p = 6.65 \pm 0.20 \text{ km s}^{-1}$. The velocity/density relation is then

$$\rho(\text{Mg m}^{-3}) = (3.50 \pm 0.2) - (3.79 \pm 0.1)/V_p(\text{km s}^{-1}) \quad (4)$$

The velocity/density relation for the porous basalt formation (equation (4)) and for the laboratory data (equation (3)) are shown in Fig. 2.

Also shown in Fig. 2 are velocities and densities estimated from downhole logging in igneous basement at a number of

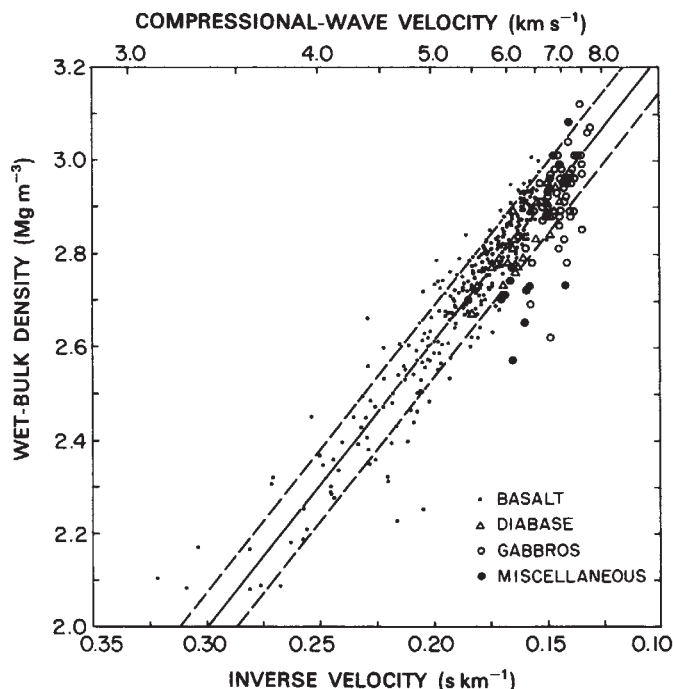


Fig. 1 Compressional-wave velocity versus wet-bulk density of basalts, diabbases and gabbroic rocks sampled from ophiolites and recovered from the sea floor by drilling. Also included are some rock types (including serpentinites and amphibolites) thought to be present but not abundant in the oceanic crust. Solid line represents least-squares fit (equation 3). Dashed lines represent standard error. The r.m.s. misfits to basalt, diabase and gabbro data taken separately range from 0.04 to 0.08 Mg m^{-3} .

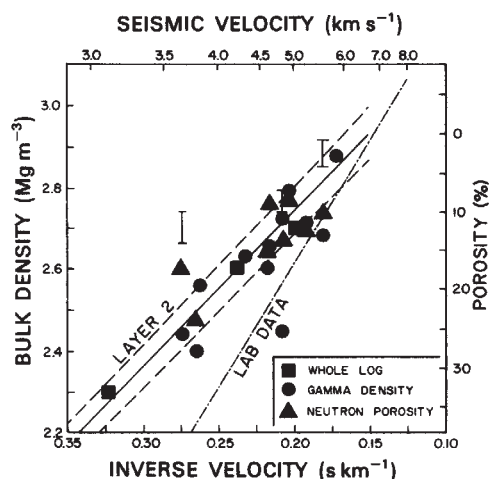


Fig. 2 Approximate velocity-density-porosity relation for a porous basalt formation (layer 2) and best-fitting relation for laboratory data. Also shown are formation velocities, densities and porosities estimated from downhole logs. Bars indicate porosities estimated from wide-spacing resistivity data²³.

DSDP sites. Some of the data are from published reports^{8,11,22}. We have made additional, preliminary analyses of logs from holes 504B (Leg 83), 556, 558 and 564. The velocities shown are averages from sonic logs. Gamma-gamma density, neutron porosity and electrical resistivity data²³ have been used to estimate densities and porosities. The neutron porosity values have been corrected for pressure and bound water content as described by Kirkpatrick⁸. Some of the data points represent averages over the entire logged interval; in other cases (such as, Leg 83, 504B) the data points represent different intervals such as layers 2a, 2b and 2c. While we must emphasize that our

interpretations of the logs are preliminary, and, therefore, subject to (presumably) minor changes, we find that the data correspond quite well with the predicted relation. The r.m.s. misfits of equation (4) to the log values of density and porosity are 0.06 Mg m^{-3} and 4%, respectively. Insofar as it does agree with available downhole logging data, equation (4) represents a valid means of estimating both the porosity and the density of layer 2 from seismic refraction data.

A particularly important point is that equation (4) predicts markedly higher densities than does the velocity-density relation based on laboratory measurements (equation (3)). For a seismic velocity of 5.0 km s^{-1} , the difference is $\sim 0.15 \text{ Mg m}^{-3}$, for a velocity typical of layer 2a, about 3.75 km s^{-1} , the difference is 0.3 Mg m^{-3} . Thus the common practice of using laboratory data to estimate formation densities from seismic velocities may lead to serious errors when the porosity of the layer is greater than the grain-boundary porosity of the rocks.

In method 3, we have accounted for large-scale porosity of layer 2a by calculating layer densities according to equation (4) when the seismic velocity is $< 6.65 \text{ km s}^{-1}$. Thus, in general, equation (4) applies to oceanic layer 2. For velocities $> 6.65 \text{ km s}^{-1}$ (layer 3), we have used equation (3).

Any of the three methods outlined above can be applied to an individual refraction survey. Because our objective is to estimate the average density of the oceanic crust, we have used several average velocity structures: three 'standard' three-layer models^{4,24,25}; the type I and II sonobouy models of Christensen and Salisbury⁴, in which layer 3 is subdivided into two units (3a and 3b); the Atlantic and Pacific multilayer models proposed by Houtz and Ewing⁶; and a four-layer model suggested by Purdy¹³ for crust 140 Myr old. The velocity structures, estimated layer densities and mean crustal densities are summarized in Table 2.

A remarkable feature of the results is that the different velocity-structure models yield such similar mean crustal densities: the range of values found by method 1 is 2.89–2.91 Mg m^{-3} , the range for method 2 is, in general, 2.84–2.86 Mg m^{-3} , and most of the densities calculated by method 3 are in the range 2.88–2.90 Mg m^{-3} . Mean crustal densities estimated by methods 1 and 3 are also remarkably similar; the average difference is $< 0.01 \text{ Mg m}^{-3}$. Method 2 yields mean densities which are systematically lower by $\sim 0.03 \text{ Mg m}^{-3}$. Having calculated the density structures by these methods, the problem is to select the best model.

Method 3 is preferable to the other methods because densities are calculated from the seismic structure, and the large-scale porosity of layer 2 is taken into account. Equation (4) used to estimate the density of layer 2 is also in good general agreement with the results of downhole logging in igneous basement (Fig. 2). Furthermore, the density structures calculated by this method are consistent with the inferred variation of physical properties with depth in the oceanic crust.

Large-scale porosity is generally restricted to the uppermost 400 m (layer 2a) of the section. Below that depth, the porosity is markedly reduced (layer 2b), and below 1,000 m (layers 2c and 3) porosity is apparently restricted to the grain boundaries^{6,11,13,23}. Consequently, calculated layer densities (Table 2) should correspond to the densities of corresponding lithologies. Calculated layer 3 densities agree with the mean density of oceanic gabbros ($2.93 \pm 0.07 \text{ Mg m}^{-3}$). The density of the lower part of layer 2 (2.88) is higher than the mean (2.85 Mg m^{-3}), but agrees with the modal value (2.88 Mg m^{-3}) for diabbases. The mean densities of different suites of oceanic basalts range from 2.79 to 2.80 Mg m^{-3} (refs 7, 10, 12). The range of calculated layer 2 mean densities (not included in Table 2) is 2.80–2.75 Mg m^{-3} . These values range downward from the mean for basalts and diabbases because of the porosity of layer 2a, but the effect of porosity on the average density of layer 2 is reduced because the porous region occupies only $\sim 20\%$ of the layer thickness.

Thus, we conclude that density structures calculated by method 3 are representative of *in situ* densities in the oceanic

crust. A reasonable mean density is $2.89 \pm 0.04 \text{ Mg m}^{-3}$. Even if we assume that 500 m of ultramafic cumulates having a density of 3.20 Mg m^{-3} should be included, the mean crustal density is only 2.91 Mg m^{-3} . The proposed existence of a low-velocity layer of partially serpentinized periodotites at the base of the crust²⁶ would not significantly affect its density because the density of periodotites containing 50% serpentine is $\sim 2.9 \text{ Mg m}^{-3}$ (ref. 27). Including a 500-m blanket of sediments having a mean density of 2.0 Mg m^{-3} reduces the density of the entire crust to $\sim 2.85 \text{ Mg m}^{-3}$.

This research was supported by Office of Naval Research Contract N-00014-80-C-0013. N. L. Carter's suggestions substantially improved the manuscript. Texas A&M Geodynamics Research Program contribution no. 44.

Received 11 May; accepted 31 July 1984.

- Watts, A. B. *J. geophys. Res.* **83**, 5989-6004 (1978).
- Molnar, P. & Atwater, T. *Earth planet. Sci. Lett.* **41**, 330-340 (1978).
- Pennington, W. D. *Tectonophysics* **102**, 377-398 (1984).
- Christensen, N. I. & Salisbury, M. H. *Rev. Geophys. Space Phys.* **13**, 57-86 (1975).
- Spudich, P. & Orcutt, J. *Rev. Geophys. Space Phys.* **18**, 627-645 (1980).
- Houtz, R. & Ewing, J. *J. geophys. Res.* **81**, 2490-2498 (1976).
- Hyndman, R. D. & Drury, M. J. *J. geophys. Res.* **81**, 4042-4052 (1976).
- Kirkpatrick, R. J. *J. geophys. Res.* **84**, 178-188 (1979).
- Anderson, R. N. *et al. Nature* **300**, 589-594 (1982).
- Hamano, Y. *Init. Rep. DSDP Legs 51, 53 Part 2*, 1457-1466 (1980).
- Salisbury, M. H. *et al. Init. Rep. DSDP Legs 51, 53 Part 2*, 1579-1597 (1980).
- Karato *et al. Init. Rep. DSDP Leg 69*, 675-682 (1981).
- Purdy, G. M. *Geophys. J. R. astr. Soc.* **72**, 115-137 (1983).
- Salisbury, M. H. & Christensen, N. I. *J. geophys. Res.* **83**, 805-817 (1978).
- Christensen, N. I. & Smewing, J. D. *J. geophys. Res.* **86**, 2545-2556 (1981).
- Christensen, N. I. & Salisbury, M. H. *Geophys. J. R. astr. Soc.* **68**, 675-688 (1982).
- Christensen, N. I. *Tectonophysics* **47**, 131-157 (1978).
- Christensen, N. I. & Salisbury, M. H. *Earth planet. Sci. Lett.* **15**, 367-375 (1972).
- Christensen, N. I. & Salisbury, M. H. *Earth planet. Sci. Lett.* **19**, 461-470 (1973).
- Fountain, D. M. *et al. Mar. Geol.* **19**, M75-M80 (1975).
- Carlson, R. L. & Wilkens, R. H. *Geodyn. Ser.* **11**, 127-136 (1983).
- Cann, J. R. & Von Herzen, R. P. *Init. Rep. DSDP Leg 69*, 281-299 (1981).
- Becker, K. *et al. Nature* **300**, 594-598 (1982).
- Raitt, R. in *The Sea Vol. 3* (ed. Hill, M. N.) 85-102 (Wiley, New York, 1963).
- Shor, G. G. in *The Sea Vol. 7* (ed. Maxwell, A.) 3-27 (Wiley, New York, 1971).
- Lewis, B. T. R. & Snodysman, W. *Nature* **266**, 340-344 (1977).
- Carlson, R. L. *et al. Earth planet. Sci. Lett.* **51**, 171-180 (1980).

Growing season precipitation from D/H ratios of Eastern White Pine

J. R. Lawrence* & J. W. C. White

Lamont-Doherty Geological Observatory of Columbia University, Palisades, New York 10964, USA

The search for climatic information from the hydrogen and oxygen isotope analysis of tree rings has produced mixed results¹⁻⁶. Systematic changes in the D/H ratio with changing climate have been demonstrated⁷, although relationships between climate change and the D/H ratio in tree rings from a single tree have been claimed, the interpretation of these data still causes controversy. Here we report that there is a strong correlation between both the D/H ratio of summer rain and tree ring cellulose from Eastern White Pine (*Pinus strobus* L.), and the amount of precipitation between May and August.

There are three major difficulties in interpreting the isotope composition of tree ring cellulose. (1) Organic matter is isotopically inhomogeneous^{4,7-11}. (2) Complex biological, biochemical and hydrological processes take place during the growth of a tree^{7,12-15}. (3) Isotope variations in precipitation from year to year at a single location have not been clearly related to temperature changes, except perhaps for those that accompany significant changes in the polar ice caps¹⁶⁻¹⁸. We have avoided the problem of the isotopic inhomogeneity of organic matter by analysing only carbon-bound hydrogen from cellulose⁷. Our studies of the complex isotopic fractionations resulting from biological, biochemical and hydrological processes in Eastern White Pine are reported elsewhere^{19,20}. They demonstrate that

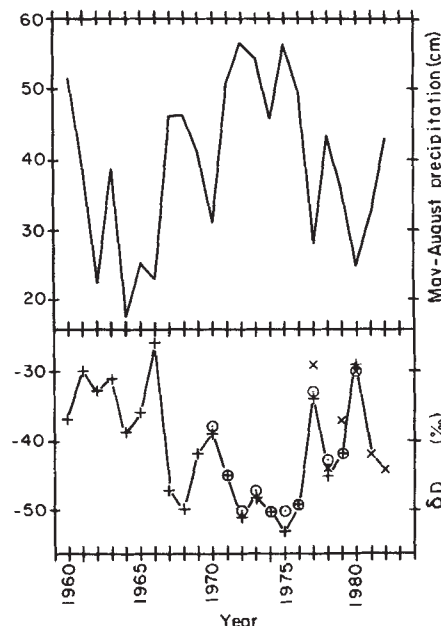


Fig. 1 Average amount of precipitation for the period May-August for five meteorological stations (Mohonk Lake, Poughkeepsie, Ellenville, Gardiner and Rosendale) plotted as a function of time. The δD values of rain and carbon-bound hydrogen from tree rings of two Eastern White Pine trees are also plotted. One tree covers the period 1960-80 (+), the other the period 1970-80 (○) and the rain from 1977-82 (x). The correlation coefficients between the δD values and the amount of May to August precipitation are -0.76 , -0.93 and -0.88 respectively.

the average hydrogen isotope composition of rain that falls during the growing season, May-August, can be found in tree rings of Eastern White Pine growing under highly restricted water supply conditions. Such trees mostly use soil moisture obtained from summer precipitation and therefore retain only a summer climate signal.

One climate signal that has been observed from stable isotopes in the precipitation from lower latitudes, is the inverse relationship between the δD and $\delta^{18}O$ of rain and the amount of precipitation. ($\delta D = [(D/H \text{ sample})/(D/H \text{ SMOW} - 1)] \times 10^3$, where SMOW is standard mean ocean water.) This 'amount effect' was observed by Dansgaard²¹ when evaluating monthly precipitation samples obtained and analysed by the International Atomic Energy Agency (IAEA). He observed that isotopic ratios were 'low in rainy months and high in months with sparse rain'. He stated: 'this "amount effect" is found all the year round at most tropical stations, and in the summer time at mid latitudes', this rain being predominantly produced by convective storms. We also detected this amount effect during the present investigation.

Our sampling location, Mohonk Lake, New York, is located ~ 130 km north of New York City on the Shawangunk ridge on the west side of the Hudson Valley near New Paltz, New York. The two Eastern White Pine trees studied are located on a talus slope where groundwater has not been detected in the tree sap²⁰. The trees are ~ 5 m high, 60 cm in diameter and > 200 -yr old.

The δD values of the rings from these trees together with the unweighted average δD value of all the individual rainfalls between May and August are plotted as a function of time in Fig. 1. One tree (no. 1) covers the period 1960-80, the other tree (no. 2) 1970-80 and the rains from 1977 to 1982. Also shown in Fig. 1 is the total amount of rain that fell between May and August of each year. These quantities represent the averages of five rain recording stations in the vicinity of Mohonk Lake.

An inverse relationship can be seen between the amount of precipitation and the isotope data (Fig. 1): years of high precipi-

* Present address: Department of Geological Sciences, University of Houston, Texas 77004, USA.

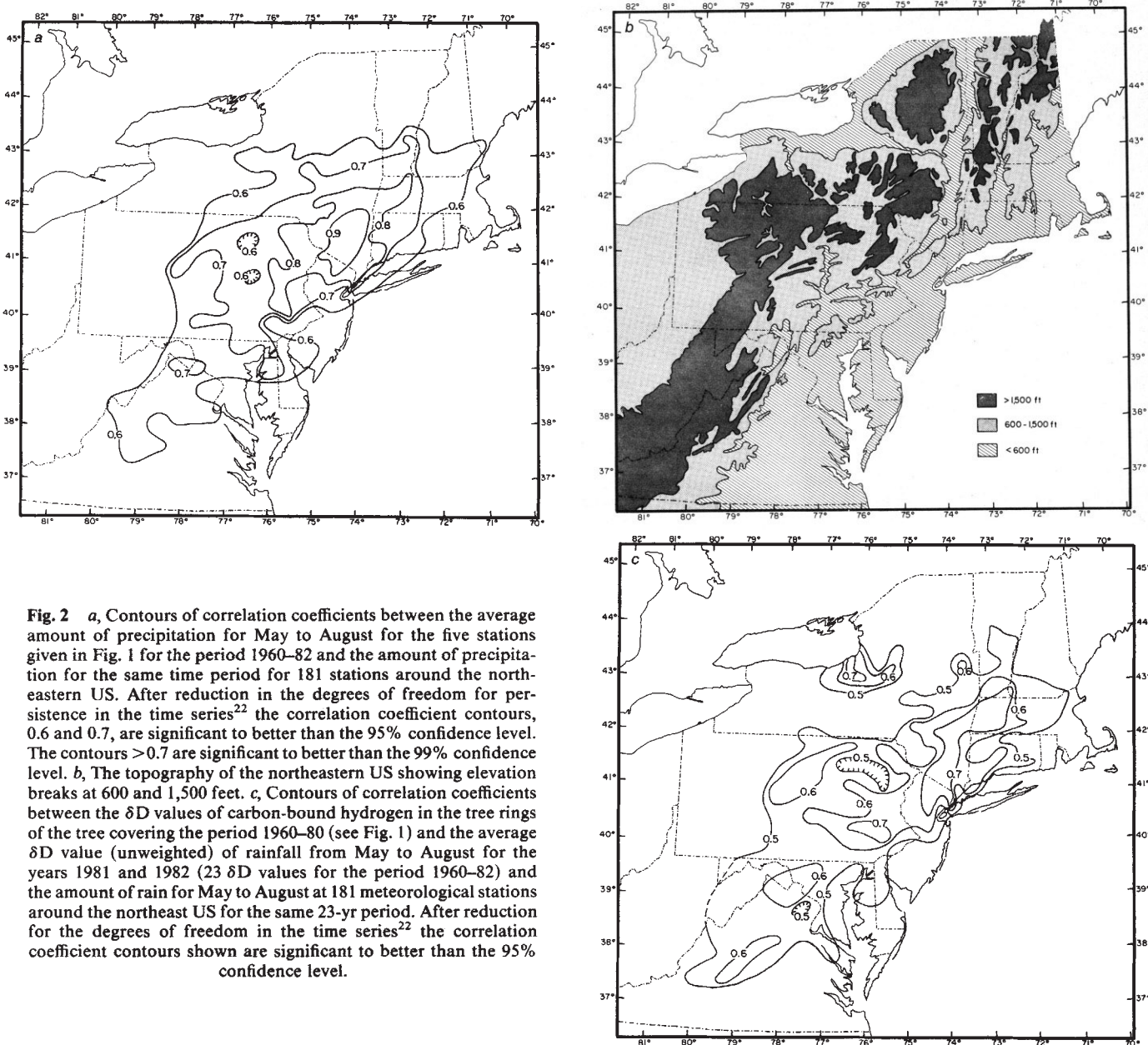


Fig. 2 *a*, Contours of correlation coefficients between the average amount of precipitation for May to August for the five stations given in Fig. 1 for the period 1960–82 and the amount of precipitation for the same time period for 181 stations around the northeastern US. After reduction in the degrees of freedom for persistence in the time series²² the correlation coefficient contours, 0.6 and 0.7, are significant to better than the 95% confidence level. The contours >0.7 are significant to better than the 99% confidence level. *b*, The topography of the northeastern US showing elevation breaks at 600 and 1,500 feet. *c*, Contours of correlation coefficients between the δD values of carbon-bound hydrogen in the tree rings of the tree covering the period 1960–80 (see Fig. 1) and the average δD value (unweighted) of rainfall from May to August for the years 1981 and 1982 (23 δD values for the period 1960–82) and the amount of rain for May to August at 181 meteorological stations around the northeast US for the same 23-yr period. After reduction for the degrees of freedom in the time series²² the correlation coefficient contours shown are significant to better than the 95% confidence level.

tation are marked by low δD values and vice versa. There is excellent agreement between the separate isotope data sets. The correlation coefficient between the mean δD value of the rain and the amount of precipitation for the 6 yr when samples were collected is -0.88 . For the tree where 11 rings were analysed, the correlation coefficient is -0.93 . For the tree covering 21 yr the correlation coefficient is -0.76 . Each of these three paired sets of data were checked for persistence (autocorrelation)²². After reduction in the number of degrees of freedom, the three correlation coefficients were found to be significant to better than the 95%, 99% and 99% confidence levels respectively. The precipitation data show that the early 1960s had low rainfall whereas the mid-1970s had high rainfall. The isotope data only provide an indication of rainfall amounts during the warmer months, but there was a severe drought during the early 1960s, and during the mid-1970s the climate was wet, as demonstrated by the high level of reservoirs at the time²³. Individual years of low rainfall during the relatively wet periods stand out—see 1980, 1977 and 1970 in Fig. 1.

The correlation coefficient between the δD of the tree rings and the amount of precipitation during the growing season applies to the immediate area around Mohonk Lake. To get a better idea of the regional extent of this isotope–precipitation

relationship, additional correlations were attempted with meteorological stations around the northeastern US. First, the average May to August precipitation amounts from the five stations in the vicinity of Mohonk Lake for the period 1960–82 were correlated with precipitation amounts for the same 4 months and 23 yr for stations around northeastern US. A total of 181 correlation coefficients were calculated. The results (shown in Fig. 2*a*) show that topography and proximity to the ocean may be important (Fig. 2*b*). Correlations drop markedly near the ocean and are generally lower over mountain areas. Lower-lying areas excluding near shore areas have higher correlation coefficients.

A similar regional correlation picture (Fig. 2*c*) was constructed using the isotope data from tree 1 for the period 1960–80 and 2 yr (1981 and 1982) of rain isotope data. The isotopic results for this 23-yr period were correlated with precipitation during May–August for the same 181 stations as used in the precipitation correlations. The patterns are very similar to those observed in Fig. 2*a*.

Our results suggest that δD values of carbon-bound hydrogen from tree rings of Eastern White Pine growing in water-restricted conditions can be used to study past summer precipitation patterns. The region covered by our studies is primarily the

Hudson Valley and bordering areas. Because the relationship between the δD values of rain and the amount of precipitation during the growing season should apply to other areas in the eastern US, isotopic studies of trees in such areas might also reveal distinctive precipitation patterns with time. Combining results from several trees in different areas would then not only give variations of growing-season precipitation for individual areas but would also give the historical distribution of precipitation in both a spatial and temporal sense. Such a picture would be useful for understanding long-term changes in atmospheric circulation particularly for the warmer months of the year.

The isotope technique for obtaining climate information from tree rings of Eastern White Pine has certain advantages including that the hydrogen isotope climate signal is derived from the precipitation itself, and thus contains information about atmos-

pheric processes. Future research will give a better picture of the causes of changes in the isotope ratios. For example, it has been shown that the isotopic composition of precipitation is influenced by the positions of storm tracks²⁴. Oxygen isotopic analyses of Eastern White Pine when combined with hydrogen isotope analyses show considerable promise in providing a relative humidity signal¹⁹; the use of both isotopes should therefore prove highly valuable. The combined use of these isotope techniques with standard ring width techniques looks even more promising.

This research was supported by grants ATM 79-18920, ATM 81-16371 and ATM 83-13954 from the Climate Dynamics Section of the NSF. We thank E. R. Cook, D. Stockton, G. Jacoby, and J. Jouzel for helpful critical comments. Lamont-Doherty Geological Observatory contribution no. 3706.

Received 3 May; accepted 13 August 1984.

- Schiegl, W. E. *Nature* **251**, 582-584 (1974).
- Wilson, A. T. & Grinsted, M. J. *Nature* **257**, 387-388 (1975).
- Libby, L. M. *et al.* *Nature* **261**, 284-288 (1976).
- Epstein, S. & Yapp, C. J. *Earth planet. Sci. Lett.* **30**, 252-261 (1977).
- Gray, J. & Thompson, P. *Nature* **262**, 481-482 (1976).
- Yapp, C. J. & Epstein, S. *Earth planet. Sci. Lett.* **34**, 333-350 (1977).
- Epstein, S., Yapp, C. J. & Hall, J. H. *Earth planet. Sci. Lett.* **30**, 241-251 (1976).
- Wilson, A. T. & Grinsted, M. J. *Earth planet. Sci. Lett.* **36**, 246-248 (1976).
- Epstein, S. & Yapp, C. *Nature* **266**, 477-478 (1977).
- Libby, L. M. & Pandolfi, L. J. *Nature* **266**, 478 (1977).
- Gray, J. & Thompson, P. *Nature* **270**, 708-709 (1977).
- Wigley, T. M. L., Gray, B. M. & Kelly, P. M. *Nature* **271**, 92-93 (1978).
- Wilson, A. T. *Nature* **271**, 93 (1978).
- Gray, J. & Thompson, P. *Nature* **271**, 93-94 (1978).
- Libby, L. M. & Pandolfi, L. J. *Nature* **271**, 94 (1978).
- Dansgaard, W. & Tauber, H. *Science* **166**, 499-502 (1969).
- Epstein, S., Sharp, R. P. & Gow, A. J. *Science* **168**, 1570-1572 (1970).
- Dansgaard, W., Johnsen, S. J., Clausen, H. B. & Langway, C. C. Jr in *Late Cenozoic Glacial Ages* (ed. Turekian, K. K.) 37-56 (Yale University, New Haven, 1971).
- White, J. W. C. thesis, Columbia Univ. (1983).
- White, J. W. C., Cook, E. R., Lawrence, J. R. & Broecker, W. S. *Geochim. cosmochim. Acta* (in the press).
- Dansgaard, W. *Tellus* **16**, 436-468 (1964).
- Dawdy, D. R. & Matalas, N. C. in *Handbook of Applied Hydrology* (ed. Chow, V. T.) 87 (McGraw-Hill, New York, 1964).
- Cook, E. R. & Jacoby, G. C. Jr *Science* **198**, 399-401 (1977).
- Lawrence, J. R., Gedzelman, S. D., White, J. W. C., Smiley, D. & Lazov, P. *Nature* **296**, 638-640 (1981).

Induction of hair growth by implantation of cultured dermal papilla cells

C. A. B. Jahoda*, K. A. Horne & R. F. Oliver

Department of Biological Sciences, University of Dundee, Dundee DD1 4HN, UK

Mammalian hairs are formed by differentiation and keratinization of cells produced in the epidermal matrix (Figs 3, 4). Using the rodent vibrissa follicle as a model¹, transplantation studies have shown that the dermal papilla, a discrete population of specialized fibroblasts, is of prime importance in the growth of hair^{2,3}. Papillae induce hair growth when implanted into follicles^{4,5} and can interact with skin epidermis to form new hair follicles⁶. When grown in culture, papilla cells display singular morphological and behavioural characteristics compared with connective tissue cells from other skin sources^{7,8}. We report here that serially cultured adult papilla cells can induce the growth of hair when implanted into follicles which otherwise would not grow hairs. This finding presents an opportunity to characterize properties distinguishing the papilla cell population from other skin fibroblasts, and, more specifically, those which control hair growth. The eventual application of this work to human hair replacement techniques can also be envisaged.

We have cultured dermal papillae from adult rats. When the cells reach confluence, they form aggregates (Fig. 1) and can be serially cultured^{7,8}. We have shown previously that discrete, freshly dissected papillae induce hair growth when implanted into the bases of follicles from which the lower halves with their original papillae have been removed⁴. Such follicles (Fig. 2) would not otherwise produce hairs^{4,9,10}. In the present studies, we used similarly prepared follicles to evaluate the inductive capacity of cultured papilla cells when implanted as pellets into the bases of follicles from which the lower halves had been removed. Skin fibroblasts of early passage numbers (1-4) were used as control implants. After 4 weeks, hairs emerged from 28 (53%) of the follicles which had received papilla cells of early passage numbers (1-3), and those follicles kept under observation produced generations of hairs in a normal cyclic fashion.

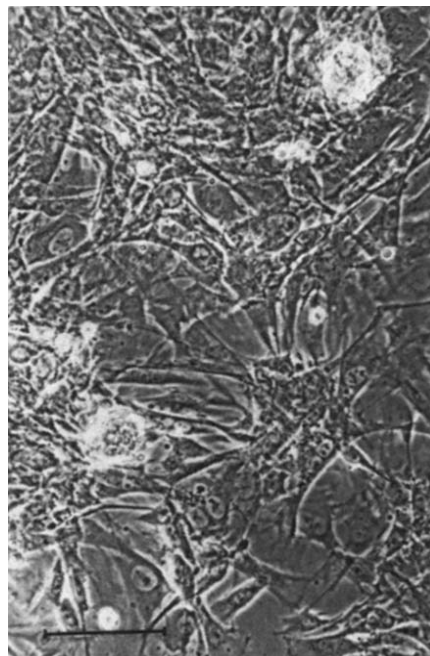


Fig. 1 Phase contrast micrograph of adult rat vibrissa dermal papilla cells on a glass substrate after 4 weeks in culture. Papilla cells, which are fibroblast-like, show a distinctive behaviour both in explant culture⁷ and after passaging⁸ and, in contrast to skin fibroblasts, tend to form clumps or aggregates on reaching confluence. Using a stereoscopic microscope, intact, epidermal cell-free dermal papillae dissected from the lower regions of vibrissa follicles, which had been excised from the upper lip of adult inbred hooded PVGC rats, were cultured as described previously^{7,8}. Between 5 and 30 papillae were cultured at any one time as explants in Cruickshank culture chambers at 37 °C in an atmosphere of 5% CO₂-95% air in Eagle's minimal essential medium (Gibco) supplemented with 15-20% fetal calf serum (Gibco). The medium was changed every 3 or 4 days and on reaching confluence the cells were serially transferred to 35-mm dishes. Adult skin fibroblasts, derived from skin explants, were similarly grown in culture. Scale bar, 100 μ m.

* Present address: Laboratoire de Zoologie, Université de Grenoble I, France.

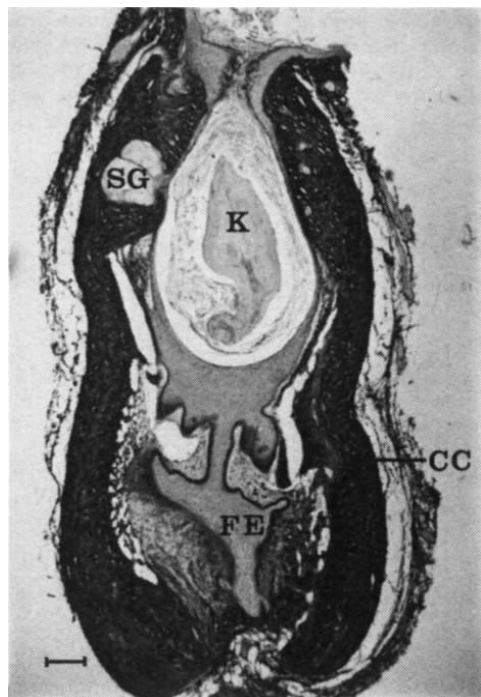


Fig. 2 Longitudinal section of a control vibrissa follicle from which approximately the lower half of the follicle has been removed. No subsequent hair growth has occurred and, near the bottom, remaining follicular epidermis (FE) has reverted to a solid column of outer root sheath cells. A loose plug of keratin (K) is visible near the top of the follicle adjacent to the sebaceous gland (SG). Vibrissa follicles are enclosed by a thick collagenous capsule (CC) the cut edge of which can be seen at the bottom. Vibrissa follicles, which are clearly arranged in a well defined pattern⁹, were exposed by reflecting the whisker pad on the upper lip and the follicles selected for operation were identified. Lengths of follicle, including the entire bulb and enclosed papilla, were removed with a transverse cut to leave half or less of the upper region of each follicle *in situ* and the pad was stitched back into position⁴. An identical procedure was used to prepare follicles to receive either cultured papilla cell or fibroblast implants. Following biopsy, follicles were serially sectioned for histological examination. Scale bar, 100 μm .

In contrast, no hairs emerged either from amputated follicles which had not received papilla cell implants (Fig. 2) or from follicles into which cultured skin fibroblasts had been implanted.

Histological examination revealed the presence of a new bulb at the site of papilla cell implantation, with an epidermal matrix organized around a discrete dermal papilla (Fig. 3). Further evidence that the new papilla originated from implanted papilla cells was obtained by autoradiographic examination of follicles where implanted papilla cells were labelled with ^3H -thymidine (Fig. 4). Thus we have shown that the original papilla cell mass can form a functional dermal papilla which interacts with follicular epidermis, resulting in the organization of a new, hair-producing bulb. By demonstrating that cultured papilla cells retain their ability to elicit hair growth, the specific properties of papilla cells which operate to modulate follicular epidermal cell activity and thus control the growth of hair can now be analysed and characterized *in vitro*. One such approach would be to compare profiles of the characteristics of papilla cells of early passage numbers (1–3) with those of later passage number cells (10–15), which appear to lose their inductive ability, and with skin fibroblasts. Work is in progress to determine the stage at which the functional capacity of papilla cells in culture is irretrievably lost.

With the recent culture of follicular epidermal cells¹¹, it will be possible also to study the influence of papilla cells (which retain some of their *in situ* traits when grown alone in culture⁸) on epidermal proliferation and differentiation. Although it has

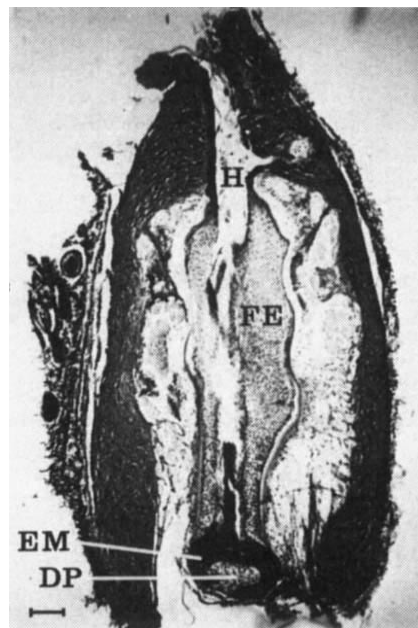


Fig. 3 The upper half of a vibrissa follicle into the base of which cultured dermal papilla cells had been implanted. Follicular epidermis (FE) has become organized around the new dermal papilla (DP) and has formed a basophilic epidermal matrix (EM) which is actively producing hair (H). Papilla cells scraped from culture dishes and divided into small pellets were inserted into the bases of the exposed follicular epidermis following removal of the lower halves of the follicles. Scale bar, 100 μm .

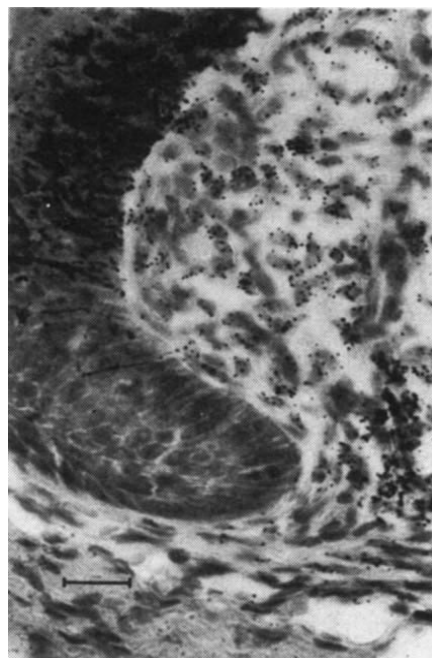


Fig. 4 Autoradiograph showing detail of the bulb region of a follicle actively growing hair 28 days after implantation of papilla cells which had been labelled in culture with a total dose of 2 μCi of ^3H -labelled thymidine. Labelling is restricted to cells in the dermal papilla and its basal stalk with no evidence of labelling in the surrounding connective tissue cells. Part of the epidermal matrix is shown on the left-hand side. Scale bar, 20 μm .

been shown that epidermal and dermal cell suspensions from newborn mouse skin are capable of reforming hair follicles when transplanted into adult animals¹², it is unclear whether at this stage of skin development all dermal cells are still capable of being involved in follicle reformation. We suggest that in adult animals, skin fibroblast cells from non-follicular dermis are

incapable of performing the same role as dermal papilla cells.

Finally, recent advances in cell culture of individual dermal and epidermal follicle components have importance for the study of human hair disease and hair loss as well as for dermal-epidermal interactions. In particular the culture of human hair dermal papilla cells¹³ means that, for example, the effects of hormones on the dermal component can now be examined. Our demonstration that dermal papilla cells which have divided can

still promote active hair growth, and our recent finding that cultured papilla cells can induce hair follicle formation when associated with non-hair-forming embryonic epidermis (C.A.B.J., unpublished data), raises the possibility of using cell culture techniques to increase dermal papilla cell numbers as part of a human hair transplant process.

We thank Mr G. Coghill for performing the autoradiography. This work was supported by MRC studentship.

Received 6 February; accepted 3 August 1984.

1. Cohen, J. *J. Embryol. exp. Morph.* **9**, 117-127 (1961).
2. Oliver, R. F. in *The Skin of Vertebrates* (eds Spearman, R. I. C. & Riley, P. A.) 199-210 (Academic, London, 1980).
3. Oliver, R. F. & Jahoda, C. A. B. in *Hair Research* (eds Orfanos, C. E., Montagna, W. & Stüttgen, G.) 18-24 (Springer, Berlin, 1981).
4. Oliver, R. F. *J. Embryol. exp. Morph.* **18**, 43-51 (1967).
5. Ibrahim, L. & Wright, E. A. *Nature* **265**, 733-734 (1977).

6. Oliver, R. F. *J. Embryol. exp. Morph.* **23**, 219-236 (1970).
7. Jahoda, C. A. B. & Oliver, R. F. *Br. J. Derm.* **105**, 623-627 (1981).
8. Jahoda, C. A. B. & Oliver, R. F. *J. Embryol. exp. Morph.* **79**, 211-224, 1984.
9. Oliver, R. F. *J. Embryol. exp. Morph.* **15**, 331-347 (1966).
10. Oliver, R. F. *J. Embryol. exp. Morph.* **16**, 231-244 (1966).
11. Weterings, P. J. J. M., Vermorken, A. J. M. & Bloemendaal, H. *Expl. Cell Res.* **139**, 439-443 (1982).
12. Worst, P. K. M., Mackenzie, I. C. & Fusenig, N. E. *Cell Tissue Res.* **225**, 65-77 (1982).
13. Messenger, A. G. *Br. J. Derm.* **110**, 685-689 (1984).

Monoclonal antibodies define differential *ras* gene expression in malignant and benign colonic diseases

A. Thor, P. Horan Hand, D. Wunderlich, A. Caruso, R. Muraro & J. Schlom

Laboratory of Tumor Immunology and Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205, USA

DNA of some human tumours can transform NIH 3T3 fibroblast cells¹⁻⁷, thus demonstrating the transforming potential of human *ras* genes (Hu-*ras*^{Ha}, Hu-*ras*^{Ki}, and Hu-*ras*^N, respectively Harvey, Kirsten and neuroblastoma *ras* genes). Only a small percentage of a given type of human carcinoma, however, scores positive in this assay system⁸⁻¹⁰. Activation of *ras* and subsequent transformation of NIH 3T3 cells are either by a point mutation in the *ras* gene¹¹⁻¹⁸ or enhanced expression of the normal, or proto-onc, *ras* gene¹⁹⁻²¹. If the transformation of a given human tumour involves the enhanced expression of the normal or cellular *ras* gene and the resulting gene product, the tumour DNA would probably score negative in the NIH 3T3 transfection assay. In human colon carcinoma, for example, lesions at position 12 of Hu-*ras*^{Ki} have been found^{8,21}. None of nine colon carcinomas obtained at biopsy, however, contain the *ras* lesion at this position, using a Hu-*ras*^{Ha} probe²²; one other colon carcinoma does appear to contain amplified proto-onc *ras*²¹, and other colon carcinomas do have increased levels of *ras* RNA²³. There are at least three explanations for these observations. Either very few colon carcinomas contain point-mutated *ras*, the lesion in the majority of colon carcinomas is at a position other than 12, or *ras* activation in many colon carcinomas involves the enhanced expression of either the point-mutated or proto-onc form of a *ras* gene. We have now used monoclonal antibodies directed against a synthetic peptide reflecting sequences of the human T24 *ras* gene product to define *ras* p21 protein expression in a spectrum of colonic disease states. Immunohistochemical analyses of individual cells within tissue sections reveal differences in *ras* p21 expression in colon carcinomas compared with normal colonic epithelium, benign colon tumours and inflammatory or dysplastic colon lesions. Our data suggest that *ras* p21 expression is correlated with depth of carcinoma within the bowel wall, and is probably a relatively late event in colon carcinogenesis.

Monoclonal antibodies directed against human tumour-associated antigens have proved extremely useful in understanding the human neoplastic process and in defining differences between human malignant and normal cell populations. Immunohistochemical techniques using formalin fixed tissue sections and monoclonal antibodies have been used to identify different cell populations within a given tumour mass, which is particularly useful in the study of human carcinomas, such as colon carcinoma.

We have developed recently a series of monoclonal antibodies that are immunoreactive with the human *ras* 21,000 molecular weight (MW) gene product, p21. The immunogen used was a synthetic peptide reflecting positions 10-17 of the Hu-*ras* protein product from the T24 bladder carcinoma²⁴. The antibodies designated RAP (for *ras* peptide) one to five²⁴, react with both the T24 Hu-*ras* peptide and the Hu-*ras*^{Ha} peptide (which differ in a single amino acid at position 12 of the p21 molecule, Val for Gly, respectively¹⁷) using solid phase radioimmunoassays. They also immunoreact with a 21,000 MW protein in both T24 bladder carcinoma cells¹⁵ and rat mammary tumours containing activated *ras*^{25,26}. We have shown also that denaturation of cell extracts or tissue sections with formalin or glutaraldehyde enhanced detection of *ras* p21 by the RAP monoclonal antibodies²⁴; this was expected, as the amino acids at positions 10-17 are predicted to sequester within the tertiary structure of the p21 molecule^{27,28}. We report here studies using the RAP antibodies to determine whether any biological correlations exist between enhanced *ras* p21 expression (either the point-mutated

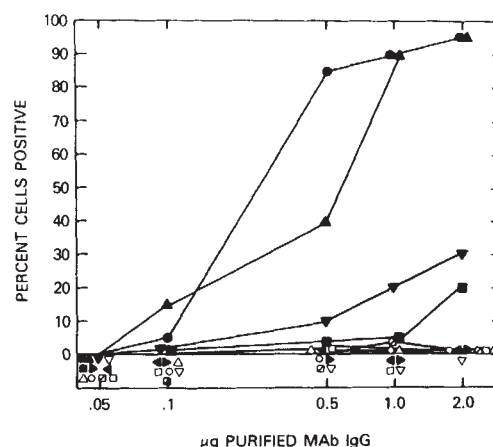


Fig. 1 Titration of purified antibody RAP-5 (in µg purified IgG per 200 µl) with formalin fixed colonic tissues. Serial sections of adenocarcinomas of the colon (solid symbols), tubulovillous adenomas (open and slashed squares), tubular adenomas (open triangles) and normal colonic epithelium (open circles) were assayed for *ras* p21 using the ABC immunoperoxidase technique^{49,50} and purified IgG⁵¹ of RAP-5. (Histologically, tubular adenomas demonstrate branching tubules embedded in the lamina propria, and tubulovillous adenomas are mixtures of tubular and villous patterns⁴³). The % of positive cells is a semiquantitative value (evaluated twice independently) determined by tumour cells scoring *ras* positive divided by the total tumour cells × 100 (for carcinomas and adenomas), and the % normal colonic epithelial cells is determined by dividing *ras* positive epithelial cells by the total number of normal colonic epithelial cells × 100. For all tissues used in this study a control IgG2a isotype identical primary antibody (UPC-10)⁵² was used at the same immunoglobulin concentrations as the RAP antibodies and no positivity was demonstrated.

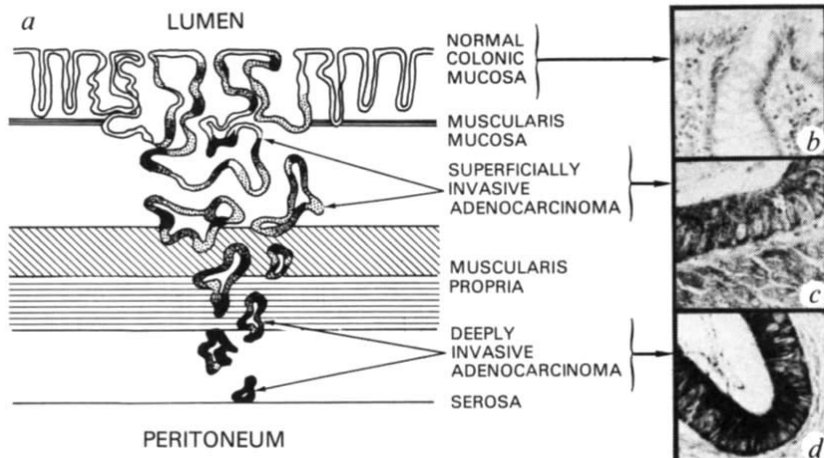


Fig. 4 Correlation between depth of colon carcinoma invasion and *ras* p21 expression. *a*, Schematic representation of a serial section of the bowel wall with the various histological layers, dysplastic and carcinomatous structures identified. Representative immunoperoxidase staining of these various structures with RAP-1 is depicted by the strong (dark) expression of *ras* p21 in deeply invasive glands versus the lesser amount of *ras* p21 expression (lighter shading) in more superficial glands. *b-d*, The different areas of a single tissue section containing all areas represented schematically in *a*. *b*, Normal colonic mucosa ($\times 220$). The epithelium and stroma are negative. There is a weak reactivity of tissue histiocytes. *c*, Superficially invasive adenocarcinoma ($\times 330$). Tumour cell heterogeneity and weak reactivity for *ras* p21 are demonstrated. *d*, Deeply invasive adenocarcinoma ($\times 220$). Subserosal tumour demonstrates intense expression of *ras* p21.

benign colon tumours, dysplastic or inflammatory lesions, or 18 normal colon specimens (all from different patients). The immunoassays reported here should be termed 'group specific' for Hu-*ras*, in that they identify determinants shared by a group of related genes (Ha, Ki, N) within a given species³⁴. The development of additional immunological reagents to 'class specific' and 'type specific' epitopes reflecting the more variable regions of this gene family, and molecular probes reflecting these variable regions, will be required to answer questions as to which tumours or cells within a given tumour mass are expressing a specific *ras* gene.

Much antigenic heterogeneity has been observed within most human carcinomas using monoclonal antibodies directed against several distinct tumour associated antigens³⁵. We find that virtually all colon carcinoma specimens also vary in the per cent of tumour cells within a given tumour mass expressing *ras* p21 (Fig. 3*a*). It is interesting that *ras* activation defined by DNA transfection occurs in only one of five melanoma cell lines derived from independent metastatic foci of a single patient³⁶. We used RAP-5 to examine five metastatic lesions from two colon carcinoma patients whose primary tumours were positive for *ras* p21 expression (40% and 80% of tumour cells, respectively). All five metastatic lesions (to ovary, liver, omentum and two lymph nodes) also displayed antigenic heterogeneity, with levels ranging from 5% to >90% of cells scoring positive for *ras* p21 expression. There are several possible explanations for this observed heterogeneity in *ras* expression. (1) The assays used may not be sensitive enough to detect low levels of *ras* expression in some cells; (2) the cell cycle may influence p21 expression (as seen in other tumour-associated antigens³⁷ and *ras*³⁸); (3) the milieu of individual tumour cells or tumour cell clusters including autocrine, paracrine and/or endocrine factors may influence *ras* expression; (4) *ras* may not be involved in the aetiology of colon cancer; its expression could be an effect of the neoplastic process in some cells or lesions. Alternatively, it may be involved in the initial transformation event, but other mechanisms of conferring immortality may come into play in the approximately 10 to 30 years that elapse between the initial cell transformation and the detection and removal of a carcinoma mass^{39,40}.

There is much evidence that many carcinomas arise from premalignant disease states such as adenomas, familial adenomatous polyposis, or ulcerative colitis⁴¹⁻⁴³. If such a multi-step process is involved in colon carcinogenesis, it would be of interest to determine where in this process enhanced *ras* expression is detected. Figure 2 shows examples of invasive carcinoma in a pseudopolyp of an ulcerative colitis patient (Fig. 2*c, e*). The dysplastic epithelial cells of the pseudopolyp here react weakly with the RAP antibody. In contrast, adjacent cells (which are more severely dysplastic and/or carcinoma *in situ*) express comparatively high levels of *ras* p21. We conclude that *ras* activation, at least in these carcinomas, may be a relatively late event in the transformation process^{10,44,45}.

The degree of malignancy (potential for metastasis) of a colon carcinoma correlates with the depth of tumour invasion through the bowel wall (from the muscularis mucosa to the serosa) (Fig. 4*a*)⁴⁶⁻⁴⁸. This invasion may occur via the infiltration of individual tumour cells through the stromal elements and musculature of the bowel wall; these cells are believed to subsequently divide and form individual tumour cell clusters with various degrees of glandular organization. We identified specimens from three colon cancer patients in which benign colonic mucosa, as well as the depth of invasion of tumour cell clusters, could be evaluated for each patient. In these samples, normal colonic mucosa are either negative or express extremely low levels ($\leq 1\%$ of cells) of Hu-*ras* p21 (Fig. 4*b*). Superficially invasive carcinoma cells show intermediate levels of Hu-*ras* p21 (Fig. 4*c*), whereas deeply invasive tumour cell clusters (Fig. 4*d*) show the highest levels of *ras* p21 in the tumour mass. Thus, in each of the three cases in which the depth of individual dysplastic and malignant tumour cell clusters could be evaluated, enhanced Hu-*ras* p21 expression detected by RAP-1 correlate with depth of tumour invasion.

Whether *ras* gene activation is actually involved in the aetiology of human colon cancer remains to be proven. The immunological studies reported here do, however, provide evidence that *ras* p21 is differentially expressed in many malignant colon carcinomas compared with benign colon tumours, dysplastic or inflammatory lesions and normal colon. Furthermore, enhanced *ras* expression appears to correlate with the depth of invasion (presumed degree of malignancy) in at least some colon carcinoma lesions, and also appears to be a relatively late event, at least within some lesions, in the multi-step process of colon carcinogenesis.

We thank K. McNaughton, A. Sloan, M. Ballard, J. Knapp, D. Poole, M. Taylor, E. Smith, P. Freeman and L. Otsby for technical assistance and Dr M. Weeks for helpful discussions.

Received 13 April; accepted 26 July 1984.

1. Pulciani, S. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2845-2849 (1982).
2. Parada, L. F., Tabin, C. J., Shih, C. & Weinberg, R. A. *Nature* **297**, 474-478 (1982).
3. Shimizu, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 2112-2116 (1983).
4. Hall, A., Marshall, C. J., Spurr, N. K. & Weiss, R. A. *Nature* **303**, 396-400 (1983).
5. Shimizu, K., Goldfarb, M., Peruchio, M. & Wigler, M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 383-387 (1983).
6. Der, C., Krontiris, T. & Cooper, G. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3637-3640 (1982).
7. Shih, C. & Weinberg, R. A. *Cell* **29**, 161-169 (1982).
8. Pulciani, S. *et al. Nature* **300**, 539-542 (1982).
9. Feig, L. A., Bast, R. A. Jr, Knapp, R. C. & Cooper, G. M. *Science* **223**, 698-700 (1984).
10. Land, H., Parada, L. F. & Weinberg, R. A. *Science* **222**, 771-777 (1983).
11. Santos, E., Tronick, S. R., Aaronson, S. A., Pulciani, S. & Barbacid, M. *Nature* **298**, 343-347 (1982).
12. Nakano, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 71-75 (1984).
13. Capon, D. J. *et al. Nature* **304**, 507-513 (1983).
14. Der, C. J. & Cooper, G. M. *Cell* **32**, 201-208 (1983).
15. Tabin, C. J. *et al. Nature* **300**, 142-148 (1982).
16. Santos, E. *et al. Science* **223**, 661-669 (1984).
17. Taparowsky, E. *et al. Nature* **300**, 762-765 (1982).
18. Reddy, E. P., Reynolds, R. K., Santos, E. & Barbacid, M. *Nature* **300**, 149-152 (1982).
19. Chang, E. H., Furth, M. E., Scolnick, E. M. & Lowy, D. R. *Nature* **297**, 479-483 (1982).
20. DeFeo, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 3328-3332 (1981).
21. McCoy, M. S. *et al. Nature* **302**, 79-81 (1983).

22. Feinberg, A. P., Vogelstein, B., Droller, M. J., Baylin, S. B. & Nelkin, B. D. *Science* **220**, 1175-1177 (1983).
23. Slamon, D. J., deKernion, J. B., Verma, I. M. & Cline, M. J. *Science* **224**, 256-262 (1984).
24. Horan Hand, P. et al. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5227-5231 (1984).
25. Sukumar, S., Notario, V., Martin-Zanca, D. & Barbacid, M. *Nature* **306**, 658-661 (1983).
26. Cho-Chung, Y. S. & Huang, F. L. in *Proc. 16th Miami Winter Symp.*, 142-143 (ICSU Press, Miami, 1984).
27. Wierenga, R. K. & Hol, W. G. J. *Nature* **302**, 842-844 (1983).
28. Pincus, M. R. et al. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5253-5257 (1983).
29. Muto, T., Bussey, H. R. J. & Morson, B. C. *Cancer* **36**, 2251-2270 (1975).
30. DeVita, V. T. Jr, Hellman, S. & Rosenberg, S. A. *Cancer: Principles and Practice of Oncology*, 647-650 (Lippincott, Philadelphia, 1982).
31. Spratt, J. S., Ackerman, L. V. & Moyer, C. A. *Ann. Surg.* **148**, 682-698 (1958).
32. Castleman, B. & Krickstein, H. I. *New Engl. J. Med.* **267**, 469-475 (1962).
33. Fung, C. H. K. & Goldman, H. *Am. J. clin. Path.* **53**, 21-25 (1970).
34. Shih, T. Y. & Weeks, M. O. *Cancer Invest.* **2**, 109-123 (1984).
35. Horan Hand, P., Nuti, M., Colcher, D. & Schlom, J. *Cancer Res.* **43**, 728-735 (1983).
36. Albino, A. P., LeStrange, R., Oliff, A. I., Furth, M. E. & Old, L. J. *Nature* **308**, 69-72 (1984).
37. Kufe, D. W. et al. *Cancer Res.* **43**, 851-857 (1983).
38. Campisi, J., Gray, H. E., Pardee, A. B., Dean, M. & Sonenshein, G. E. *Cell* **36**, 241-247 (1984).
39. Spratt, J. S. & Ackerman, L. V. *Am. Surg.* **27**, 23-28 (1961).
40. Welin, S., Youker, J. & Spratt, J. S. *Am. J. Roentgenol., Radium Ther. nucl. Med.* **90**, 673-687 (1963).
41. Bussey, H. R. *Familial Polyposis Coli* (Johns Hopkins University Press, Baltimore, 1975).
42. Edwards, F. C. & Truelove, S. C. *Gut* **5**, 15-22 (1964).
43. DeVita, V. T., Jr, Hellman, S. & Rosenberg, S. A. *Cancer: Principles and Practice of Oncology*, 674-675 (Lippincott, Philadelphia, 1982).
44. Land, H., Parada, L. F. & Weinberg, R. A. *Nature* **304**, 596-602 (1983).
45. Needleman, S. W., Yuasa, Y., Srivastava, S. & Aaronson, S. A. *Science* **222**, 173-175 (1983).
46. Dukes, C. E. *J. Path. Bact.* **50**, 527-539 (1940).
47. Cass, A. W., Million, R. R. & Pfatt, W. W. *Cancer* **37**, 2861-2865 (1976).
48. Molson, B. C. & Dawson, I. M. P. *Gastrointestinal Pathology*, 664-665 (Blackwell, Oxford, 1979).
49. Nuti, M. et al. *Int. J. Cancer* **29**, 539-545 (1982).
50. Hsu, S.-M., Raine, L. & Fanger, H. J. *Histochem. Cytol.* **29**, 577-580 (1981).
51. Colcher, D. et al. *Cancer Res.* **43**, 736-742 (1983).
52. Potter, M. *Physiol. Rev.* **631**-719 (1972).

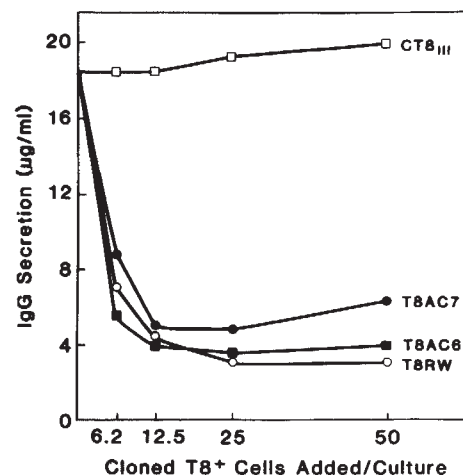


Fig. 1 Regulatory activities of T8+ T-cell clones. Human T8+ T-cells T8AC6 and T8AC7 were derived by stimulating PBMC of a healthy atopic donor with irradiated autologous antigen/MHC restricted T4+ inducer RW17C and cloning in a semi-solid system as described previously⁸. T8RW was obtained by an identical technique from RWAGE stimulated PBMC of the same donor. Individual colonies were expanded in liquid culture by restimulating with irradiated (5,000 rad) autologous PBMC and IL-2 containing supernatants. Further cloning was performed ($\times 2$) by limiting dilution technique at 0.3 cells per well. Clones were maintained in IL-2 and also placed on irradiated autologous PMBC on a feeder layer every 3 weeks. CT8III is an extensively characterized alloreactive T8+ T-cell clone derived from the same donor and was utilized as a control¹. Varying numbers of T8+ cloned cells (6.2×10^3 to 50×10^3) were added to a culture consisting of 50×10^3 autologous B cells (E rosette negative PBMC treated with anti-T4, anti-T8, anti-Mo1 and rabbit complement)¹⁶, 2.5×10^3 macrophages, 50×10^3 RW17C and the specific antigen RWAGE (final concentration: $5 \mu\text{g ml}^{-1}$) (Worthington Diagnostics³). The cultures were performed in a total volume of 0.2 ml in 96 well round-bottomed microtitre plates (Costar) in RPMI 1640 (Gibco) containing 10% FCS (M.A. Bioproducts). After 7 days' incubation at 37°C in a humidified atmosphere containing 7% CO₂ in air, the wells were harvested by pooling the cells and supernatants from triplicate wells and storing at -20°C prior to assay of IgG synthesis. Solid phase RIA was performed as previously described¹⁸.

T3-Ti receptor triggering of T8+ suppressor T cells leads to unresponsiveness to interleukin-2

Armand Bensussan, Oreste Acuto, Rebecca E. Hussey, Claudio Milanese & Ellis L. Reinherz

Division of Tumour Immunology, Dana-Farber Cancer Institute and the Departments of Medicine and Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA

T3-associated disulphide linked heterodimers (Ti_n) comprised of clonally unique α -chains of molecular weight (MW) 49,000-54,000 and β chains of MW 43,000 have been identified as the antigen receptors on human cytotoxic effector and inducer T-lymphocytes¹⁻³. Crosslinking of Ti molecules by either the appropriate nominal antigen/MHC specificity or anti-clonotypic monoclonal antibody results in clonal expansion of such cells via induction of IL-2 receptor expression, endogenous IL-2 release and IL-2-IL-2 receptor interaction⁴. To determine whether analogous antigen receptor molecules and autocrine growth mechanisms are utilized by suppressor T-cells⁵⁻⁷, we produced an anti-clonotypic monoclonal antibody against a non-cytotoxic T8+ suppressor T-cell, T8AC6, which defines a T3-associated disulphide-linked heterodimer of similar molecular weight to the above clonotypes. We find that T3-Ti triggering of suppressor clones (T8AC6, T8AC7 or T8RW) does not result in IL-2 production or T-cell proliferation and in contrast to inducer clones, also leads to a transient IL-2 unresponsive state. We suggest that such T3-Ti receptor mediated autoregulation of suppressor T-cell growth is necessary in the facilitation of initial inducer T-cell activation following antigenic perturbation.

We derived human suppressor T-cells T8AC6 and T8AC7 by stimulating autologous peripheral blood mononuclear cells (PBMC) of a healthy ragweed antigen (RWAGE) atopic donor with an MHC restricted/RWAGE-antigen specific inducer T-cell termed RW17C³ and cloned in a two-layer agar system as previously described⁸. This strategy was employed because earlier studies indicated that such stimulation results in both autoreactive T4+ amplifier T-cells and noncytotoxic T8+ clones⁸. A third suppressor clone (T8RW) was obtained from direct stimulation of PBMC with RWAGE at the end of the fall hay fever season. We propagated these T-cell clones *in vitro*

with IL-2 and irradiated autologous PBMC for more than 14 months, maintaining stable phenotype (T1+T3+T4-T8+T11+T12+) and function. To determine whether the T8+ clones were capable of suppressing the secretion of immunoglobulin by a combination of B-cells and RW17C T-cells stimulated with RWAGE, we added increasing numbers of T8AC6, T8AC7 or T8RW to the autologous T-B-cell mixture consisting of 5×10^4 T4+RW17C T-cells in the presence of RWAGE. This mixture produces $18 \mu\text{g ml}^{-1}$ of IgG without additional cells (Fig. 1). In the absence of RW17C, B-cells with or without RWAGE produces less than 200 ng of IgG per ml (data not shown). Addition of increasing numbers of T8+ cells results in marked suppression of IgG secretion. For example, when as little as 6.2×10^3 cloned T8-cells were added, IgG secretion is reduced to 6-9 $\mu\text{g ml}^{-1}$, a reduction of more than 50%, and is reduced further when 12.5 - 50×10^3 T8+ T-cells were added.

To eliminate the possibility that the inhibition of immunoglobulin production is due to an unspecific crowding effect of cells in the culture system, an autologous alloreactive cytotoxic T-cell clone, CT8III, was added to the mixture of RW17C T-cells and B-cells. In contrast to the effect of the three other clones (T8AC6, T8AC7 and T8RW), increasing CT8III concentration has no effect on IgG secretion by RW17C and B-cells (Fig. 1). We conclude that the suppressive activity of T8+ clones in this system is not mediating via a trivial crowding mechanism.

We then investigated whether clonotypic structures on T8+ suppressor T-cell clones were similar to those identified on T4+ inducer clones and class I and class II specific CTL clones. We

Table 1 Influence of various monoclonal antibodies on IL-2 proliferative responses of clones T8AC6 and T4AC3

		Pretreatment of clone T8AC6					
Stimulus	Medium	Anti-T3	Anti-Ti ₅	Anti-Ti ₆	Anti-T4	Anti-T8	
Medium	480	398	456	506	511	413	
IL-2	27,312	5,301	6,205	29,312	28,049	28,506	
		Pretreatment of clone T4AC3					
Stimulus	Medium	Anti-T3	Anti-Ti ₅	Anti-Ti ₆	Anti-T4	Anti-T8	
Medium	140	153	190	192	201	141	
IL-2	16,912	26,281	17,412	27,266	18,733	17,854	

3×10^4 T8AC6 or T4AC3 cells were incubated in round-bottomed microtitre wells (Costar) with either media or monoclonal antibody (final dilution 1:1,000 in culture medium). 30 h later, purified 0.5 IL-2 U ml⁻¹ (provided by Dr K. A. Smith, Dartmouth Medical School) was added⁴. After 24 h at 37 °C in humidified atmosphere containing 7% CO₂ in air, individual wells were pulsed with 0.2 µCi of ³H-TR (Schwarz-Mann), collected 18 h later on a MASH II apparatus (M. J. Bioproducts) and ³H-TdR incorporation was then measured in a β scintillation spectrometer. Each value represents the mean of triplicate samples with standard deviations in all experiments <15%. Anti-Ti₅ and anti-Ti₆ are the anti-clonotypic antibodies directed against T8AC6 and T4AC3, respectively. These IgG1 subclass monoclonal antibodies were also purified from malignant ascites with protein A Sepharose (Pharmacia) and coupled to swollen CnBr activated Sepharose 4B beads (Pharmacia) at 5 mg ml⁻¹ for induction of clonal proliferation and IL-2 production⁴. Proliferative responses of 3×10^4 T8AC6 or 3×10^4 T4AC3 were as follows: (i) T8AC6: + medium, 392 c.p.m.; + Sepharose-linked anti-T3, 282 c.p.m.; + Sepharose-linked anti-Ti₅, 314 c.p.m.; + Sepharose-linked anti-Ti₆, 416 c.p.m.; + Sepharose-linked anti-T4, 402 c.p.m.; + Sepharose-linked anti-T8, 511 c.p.m. (ii) T4AC3: + medium, 153 c.p.m.; + Sepharose-linked anti-T3, 19,373 c.p.m.; + Sepharose-linked anti-Ti₅, 198 c.p.m.; + Sepharose-linked anti-Ti₆, 15,219 c.p.m.; + Sepharose-linked anti-T4, 202 c.p.m.; + Sepharose-linked anti-T8, 186 c.p.m. To quantitate the amount of IL-2 secreted by T8AC6, T8AC7, T8RW, and T4AC3 after stimulation with Sepharose-linked anti-T3, a previously described murine T-cell proliferative assay was employed⁴. Specifically, 10^5 T-lymphocytes were incubated overnight at 37 °C with Sepharose-linked anti-T3 in a final volume of 0.5 ml. The supernatants were then collected, passed through 0.45 µ filters and added 1:1 (v/v) to 5×10^4 murine PHA T-lymphoblasts. After 18 h of culture, individual wells were pulsed for 24 h with 0.2 µCi of ³H-TdR, mashed and counted. A standard curve obtained with serial dilutions of purified human IL-2 was used to quantitate the amounts of lymphokine produced by the various T-cell clones. Supernatants generated from T8AC6, <0.01 U ml⁻¹; T8AC7, <0.01 U ml⁻¹; T8RW, <0.01 U ml⁻¹; T4AC3, >0.20 U ml⁻¹.

developed a series of monoclonal antibodies directed at the T8AC6 suppressor T-cell clone by standard fusion techniques⁹, and selected hybridoma cultures containing antibody reactive with T8AC6 but unreactive with Laz 509 (an autologous EBV transformed line). Supernatant from one individual hybridoma of approximately 800 tested contained antibody (anti-Ti₅, IgG1 murine isotype subclass) which did not react with resting or activated T-cells or a series of additional T8+ suppressor clones derived from the same individual ($n = 15$). This hybridoma was cloned and recloned by limiting dilution in the presence of feeder cells; malignant ascites were then developed.

Solubilized membrane preparations were obtained from the ¹²⁵I labelled T8AC6 clone in order to characterize the surface structure recognized by anti-Ti₅. SDS-PAGE electrophoresis of the molecule precipitated by anti-Ti₅ from ¹²⁵I labelled-T8AC6 in reducing conditions shows two bands of a 49,000 and 43,000 MW corresponding to the α- and β-chain respectively (Fig. 2, lane a). In nonreducing conditions this structure appears as a single band of 90,000 MW (Fig. 2, lane c), which demonstrates that the Ti₅ structure is a disulphide-linked heterodimer. The 90,000, 49,000 and 43,000 molecular weight species are not present in control immunoprecipitates with an irrelevant anticonotype, anti-Ti₆ (Fig. 2, lanes b and d). Anti-Ti₅ does not precipitate material from a second T8+ suppressor clone T8AC7, derived from the same donor, thus confirming its non-crossreactive anticonotypic specificity (data not shown).

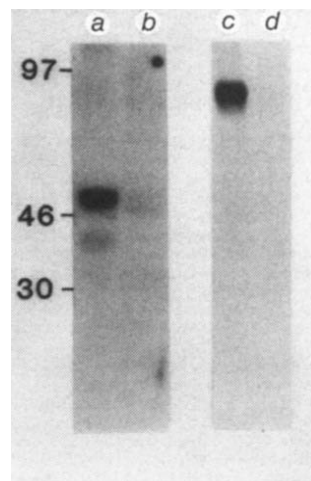


Fig. 2 SDS-PAGE of anti-Ti₅ immunoprecipitates from surface ¹²⁵I-labelled T8AC6 cells. Cells were surface labelled with ¹²⁵I (NEN) using a standard lactoperoxidase method¹⁰. Briefly, 2×10^7 viable, cells were washed with RPMI 1640, resuspended in PBS, and labelled with 1 µCi ¹²⁵I for 15 min at room temperature, followed by the addition of 100 µl NaI (1 mol l⁻¹). Labelled cells were then harvested and washed three times in cold RPMI 1640. Cell lysate was prepared by resuspending the cell pellet in 300 µl RIPA buffer, pH 7.5, containing 1% Triton X-100, phenylmethylsulphonyl fluoride, and trypsin inhibitor and agitated for 40 min at 4 °C. Cell lysates were stored at -70 °C before use. Immune precipitations were carried out using preformed complexes of rabbit anti-mouse Ig and specific monoclonal antibody¹⁹. They were then washed 4 times with RIPA buffer as above and dissolved in SDS-polyacrylamide gel electrophoresis (SDS-PAGE) sample buffer and loaded onto separate slots on a 10% polyacrylamide slab gel. Identical aliquots of immune precipitates were run in reduced and nonreduced conditions. Gels were dried and radio-labelled precipitates were visualized using standard methods. a, Anti-Ti₅, reduced; b, negative control complex (anti-Ti₆, an irrelevant anticonotypic antibody), reduced; c, anti-Ti₅, non-reduced; d, negative control complex (anti-Ti₆), non-reduced.

Previous studies with cytotoxic T cell clones indicate that the Ti structure is linked to the monomeric 20,000 and 25,000 MW T3 molecules and comodulates with them¹⁰; we have therefore attempted to determine whether a similar noncovalent association exists between the Ti molecules on suppressor cells and T3. We preincubated T8AC6 cells with anti-T3 for 24 h at 37 °C and then washed them to remove free monoclonal antibodies. We subsequently analysed cell reactivity by indirect immunofluorescence (Epics V cell sorter) with a panel of monoclonal antibodies. Before modulation, we found that all T8AC6 cells are reactive with anti-T3 and anti-Ti₅ (Fig. 3). After modulation with anti-T3, however, T3 expression is reduced by 95%. Anti-T3 induced modulation also results in removal of the Ti₅ molecule, indicated by loss of anti-Ti₅ reactivity. This is not a non-specific effect, as we observe that T8 surface expression is not influenced by this process and IL-2 receptor expression (shown by anti-IL-2 R, a monoclonal antibody directed at the IL-2 receptor and with an identical distribution to anti-Tac) is enhanced. Our findings imply that suppressor T-cells have antigen receptor structures similar to those found on inducer and cytotoxic T-cells (T3-Ti molecular complexes).

We show here that the outcome of a given immune response is dependent on the relative number of T8+ suppressor and T4+ inducer T-cells (see also ref. 11). It is thus important to determine the signals required for T8+ suppressor T-cell growth. Earlier studies have shown that soluble or Sepharose bound anti-Ti (or anti-T3) monoclonal antibodies, like physiological ligand, enhanced proliferative responses of inducer and cytotoxic T-cells to purified IL-2 by inducing an increase in their surface IL-2 receptor expression⁴. Moreover, Sepharose-bound anti-Ti (or anti-T3) or physiological ligand triggers

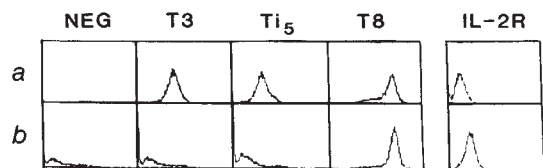


Fig. 3 The Ti structure of the T8+ T-cell suppressor clone T8AC6 comodulates with T3. T8AC6 cells were preincubated with media (a) or the monoclonal antibody anti-T3 (b) (final dilution 1:300) for 24 h at 37 °C and subsequently washed several times prior to analysis. Phenotypic analysis was performed by indirect immunofluorescence with fluorescence conjugated goat anti-mouse F(ab')₂ IgG. Samples were analysed with monoclonal antibodies (see designations above indicated histogram) on an Epics V cell sorter¹⁰. 10,000 cells were analysed in each sample. Each histogram displays the number of cells (ordinate) versus the intensity of fluorescence (abscissa) expressed on a logarithmic scale. The negative control used to determine background fluorescence was a 1:300 dilution of ascites derived from a nonreactive hybridoma (neg).

endogenous clonal IL-2 production in the absence of macrophages or feeder cells, leading subsequently to T-cell proliferation. This suggests that induction of IL-2 receptor expression, but not IL-2 release occurs with minimal T3-Ti receptor crosslinking and that antigen-induced proliferation of these cells is mediated through an autocrine pathway involving endogenous IL-2 production, release and subsequent binding to IL-2 receptors. We have performed similar experiments using T8+ suppressor T-cells. The representative T8+ suppressor clone (T8AC6) proliferates as well as the T4+ inducer clone T4AC3 to purified IL-2 (measured by ³H-TdR incorporation) (Table 1). In the presence of soluble anti-T3 antibody, however, the IL-2 response of T8AC6 was virtually abolished (5,301 vs. 27,312 cpm). Similarly, soluble anti-Ti₅ greatly reduces IL-2 responsiveness of T8AC6. An irrelevant anti-clonotype, (anti-Ti₆), or anti-T4 or anti-T8 antibodies have no effect on subsequent IL-2 responsiveness by T8AC6, and anti-Ti₅ has no effect on the response of the T4AC3 clone to IL-2. Hence the effects on T8 suppressor clones are specific and non-toxic. Furthermore, although the appropriate anti-clonotype and anti-T3 antibody inhibit IL-2 responsiveness of T8+ suppressor cells, they also enhance the IL-2 response of T4AC3 T4 inducer cells as previously observed (16,912 to 26,281–27,266 c.p.m.)¹⁰. This enhancement was observed at all anti-T3 and anti-Ti₆ antibody concentrations tested (10⁻²–10⁻⁴ dilutions), and in either 0.1, 0.5 or 1 U ml⁻¹ of IL-2. The inhibition of IL-2 responsiveness following soluble anti-T3 or anti-Ti preincubation of suppressor T-cells is not secondary to a failure of IL-2 receptor expression (Fig. 3).

One possible explanation of the inhibition is that such T3-Ti triggering results in an autoregulatory antagonist functional activity, in which case even T3-Ti receptor crosslinking with Sepharose-bound antibody would not initiate cell growth. We find that such Sepharose-bound anti-T3 and appropriate anti-Ti antibodies induce T4AC3 but not T8AC6 proliferation (Table 1). These antibodies stimulate IL-2 production from T4AC3 but not T8AC6, T8AC7 or T8RW. As crosslinking of T3/Ti molecules on suppressor T-cells does not result in autologous IL-2 production and clonal proliferation, it is not surprising that the same T8+ suppressor T-cells are unable to proliferate to the irradiated RW17C inducer T-cell utilized in the initial cloning procedure, autologous macrophages or B-cells (with or without RWAGE). The absence of a measurable proliferative response makes determination of the specificity and the target of an individual suppressor clone difficult. We plan experiments on regulation of immunoglobulin production with suppressor clones and isolated autologous and allogeneic lymphoid populations to address this issue.

Our results described here suggest that suppressor T-cells proliferate to IL-2 but that triggering of the T3-Ti complex on suppressor T-cells (with or without crosslinking) induces a state of IL-2 hyporesponsiveness. We also find that this refractory

period is transient, persisting for only three to four days. Thus, the growth response of anti-Ti (or anti-T3)-treated T8AC6 compared to anti-T8-treated (control) T8AC6 cells is reduced by 80% at 24 h, 60% at 48 h, 30% at 72 h and 0% at 96 h. Such autoregulation of suppressor T-cell growth^{12,13} may be necessary to facilitate inducer T-cell activation following initial antigenic perturbation. Once antigen is eliminated and T3-Ti stimulation abates, inducer T-cell-derived IL-2 can mediate clonal T8+ suppressor cell growth. Persistent amplification of the immune response would result in further T8+ suppressor cell expansion and hence appropriate termination of induction. Our future experiments will attempt to determine the targets and mechanism of suppression, and the relationship of the T3-Ti complex to any putative suppressor effector molecules which abolish immunoglobulin production in this or other unrelated systems^{6,14,15}. We will also attempt to determine whether other types of suppressor T-cells are governed by similar mechanisms of growth regulation^{16,17}. We thank Mr John Daley and Mr Herb Levine for technical assistance with the Epics V cell sorter. This work was supported by NIH grants R01 NS17182 and R01 AI19087-01. A.B. is a recipient of a research support grant from the French Medical Research Foundation.

Note added in proof: Monoclonal antibodies and heteroantiseras have now been produced to clonotypes on T8RW and T8AC7 and precipitate characteristic disulphide-linked heterodimers.

Received 1 May; accepted 16 July 1984.

- Meuer, S. C. *et al.* *Nature* **303**, 808–810 (1983).
- Acuto, O. *et al.* *J. exp. Med.* **158**, 1368–1373 (1983).
- Meuer, S. C. *et al.* *Science* **222**, 1239–1242 (1983).
- Meuer, S. C. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **81**, 1509–1513 (1984).
- Cantor, H. F. *et al.* *J. exp. Med.* **143**, 1391–1401 (1976).
- Fresno, M. *et al.* *J. exp. Med.* **153**, 1246–1259 (1981).
- Sy, M. S. *et al.* *J. exp. Med.* **153**, 1415–1421 (1981).
- Bensussan, A. *et al.* *J. exp. Med.* **159**, 559–576 (1984).
- Kohler, G. & Milstein, C. *Nature* **256**, 495–497 (1975).
- Meuer, S. C. *et al.* *J. exp. Med.* **157**, 705–719 (1983).
- Reinherz, E. L. *et al.* *Eur. J. Immun.* **10**, 570–572 (1980).
- Gershon, R. K. *et al.* *J. exp. Med.* **153**, 1533–1546 (1981).
- Green, J. R. *et al.* *Eur. J. Immun.* **11**, 973–986 (1981).
- Tanaguichi, M., Takei, I. & Tada, T. *Nature* **283**, 227–228 (1980).
- Kapp, J. A., Sorensen, C. M. & Pierce, C. W. *Prog. Immun.* **5**, 871–880 (1983).
- Benacerraf, B. & Germain, R. N. *Immun. Rev.* **38**, 70–119 (1978).
- Lamb, J. R. & Feldman, M. *Nature* **300**, 456–458 (1982).
- Reinherz, E. L. *et al.* *J. Immun.* **128**, 463–468 (1982).
- van Aghoven, A. C. *et al.* *Eur. J. Immun.* **11**, 18–21 (1981).

A Ca-dependent Cl⁻ conductance in cultured mouse spinal neurones

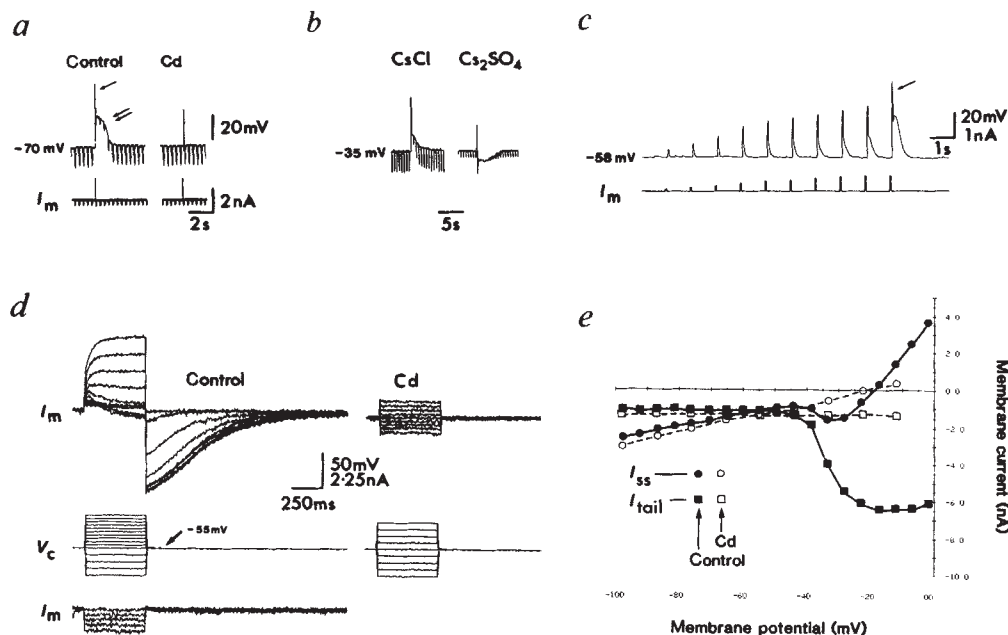
David G. Owen, Menahem Segal & Jeffery L. Barker

Laboratory of Neurophysiology, National Institute of Neurological and Communicative Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20205, USA

Long-lasting conductance changes triggered either by brief (milli-second) electrical stimuli and/or entry of calcium ions have been observed in a variety of excitable tissues^{1–4}. The electrical consequences of these events depend on the ion conductance affected and on the ion concentration gradient across the membrane, while the long-lasting nature of the change sustains the cell at either sub- or supra-threshold levels for activation of regenerative action potentials. We report here that many cultured mouse spinal neurones exhibit a voltage-activated chloride conductance that lasts for seconds and is dependent on extracellular calcium, [Ca²⁺]_o. This conductance may repolarize and stabilize the cell at a level subthreshold for generating action potentials, thus complementing the functional roles of Ca-dependent K⁺ conductances^{5–8}.

Neurones were cultured from the embryonic mouse spinal cord according to methods published previously⁹. After 3–4 weeks in culture, we studied the cells at 23 °C with intracellular recording and two-electrode voltage-clamp techniques (Fig. 1). Most of our experiments used CsCl-filled microelectrodes and bathing media containing tetraethylammonium (TEA) in order

Fig. 1 A slow Ca^{2+} - and Cl^- -dependent conductance is present in cultured mouse spinal neurones. **a**, Records made using a KCl-filled electrode. Left trace: A 50-ms depolarizing current step evokes a Ca spike (single arrow) and long-lasting ('slow') afterdepolarization (double arrow). Voltage deflections in response to constant-amplitude hyperpolarizing current pulses (I_m) reveal an increase in membrane conductance during the afterdepolarization. Right trace: Both events are abolished when 0.2 mM Cd^{2+} is applied by pressure ejection. **b**, Voltage responses to 50-ms depolarizing current steps at a holding potential (maintained with d.c. current injection) of -35 mV. A depolarizing response is recorded with a CsCl-filled electrode (left trace) while a hyperpolarizing response is triggered with Cs_2SO_4 -filled electrodes (right trace). Downward voltage deflections are in response to hyperpolarizing constant current injections and indicate sizeable increases in membrane conductance during both responses. **c**, Voltage responses (upper trace) recorded with a CsCl-filled electrode in response to depolarizing constant current steps of increasing amplitude (I_m). Activation of the long-lasting depolarizing response occurs in a graded manner and precedes the generation of a Ca spike (indicated by single arrow). **d**, Voltage-clamp analysis of the same cell shown in **c** using two CsCl-filled electrodes. Membrane currents (I_m) in response to voltage commands (V_c) in normal bathing medium (left) and in the presence of Cd (right). For clarity, control current records (left) are separated according to the polarity of the commands, with responses to hyperpolarizing command voltage steps shown in the lower traces and responses to depolarizing steps in the upper traces. Voltage commands and data acquisition were controlled with a PDP-11/23 computer. Current and voltage were digitized at 250 Hz and for display purposes individual data points are connected by lines. **e**, Current-voltage (I - V) relationships for data obtained from **d**. Currents were measured at the end of the 500-ms voltage steps (steady-state currents, I_{ss}), and immediately after returning to the holding potential (I_{tail}), in the presence and absence of 0.1 mM Cd^{2+} . In all these experiments, microelectrodes contained either 3 M KCl, 3 M CsCl or 3 M Cs_2SO_4 . The bathing medium consisted of (mM): 121 NaCl, 2.5 KCl, 4 MgCl_2 , 4 CaCl_2 , 5 HEPES, 10 glucose, 20 TEA-Cl and 0.5-1.0 μM TTX.



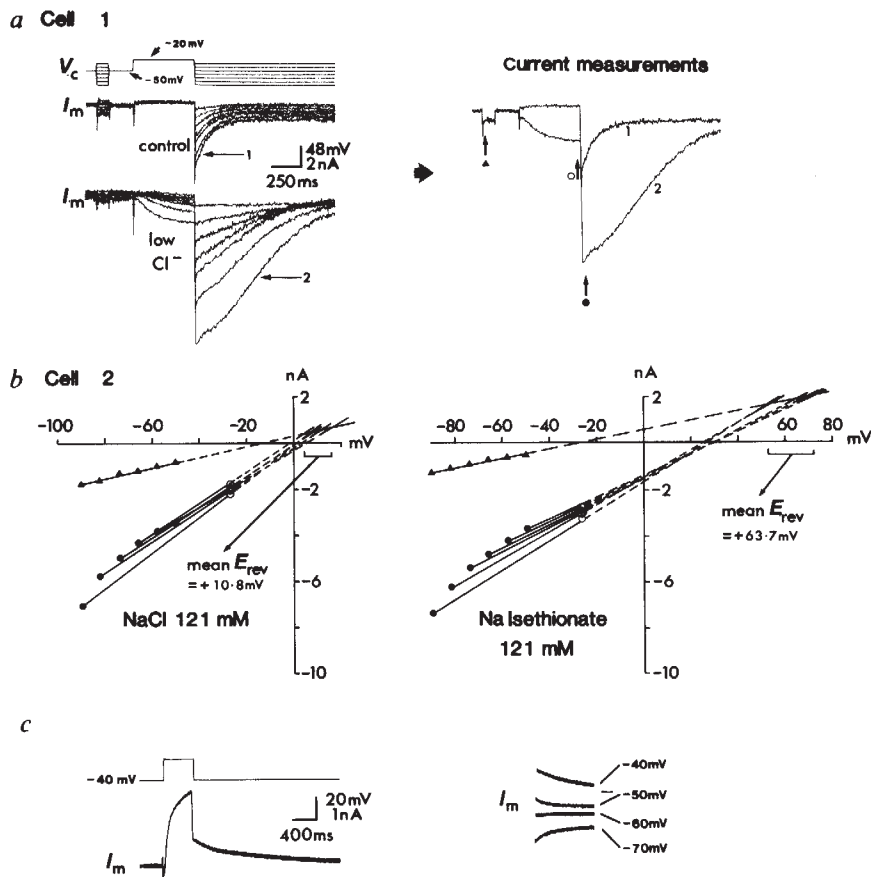
to block all outward potassium conductances;^{10,11}. In these conditions, 27 of 33 cells expressed the conductance described below, whereas with KCl-filled electrodes the conductance was detected less frequently (10 of 43 cells).

The conductance response was routinely elicited with brief depolarizing current steps to potentials more depolarized than -45 mV. When recording with KCl-filled electrodes, the response was usually associated with a 15-30 mV depolarization (from resting potentials of -60 to -70 mV) lasting a second or more and usually preceded by a rapid (about 100 ms) Ca^{2+} spike (Fig. 1a). Both the Ca^{2+} spike and the slow after-depolarization were eliminated by local application of cadmium (Fig. 1a) or cobalt ions. The polarity of the voltage change associated with the conductance response was sensitive to the counteranion in the microelectrode. A slow depolarizing conductance was consistently triggered in cells recorded with CsCl-filled electrodes, whereas a long-lasting hyperpolarizing response could be triggered in cells when recording with Cs_2SO_4 -filled electrodes (Fig. 1b). The Cl^- gradient across the membrane is therefore important in determining the polarity of the response. A Ca -dependent action potential was not a prerequisite for triggering the conductance response, as activation of the slow depolarization frequently occurred in a graded manner and at membrane potentials negative to the threshold for generating a Ca -dependent action potential (Fig. 1c). The steady-state current-voltage (I - V) relationship (where the steady-state current was measured at the end of each 500-ms voltage step) of cells bathed in TEA and tetrodotoxin (TTX) and voltage-clamped with CsCl-filled electrodes was nonlinear over the range of membrane potentials -45 mV to 0 mV (Fig. 1d, e). Following voltage steps to potentials more depolarized than -45 mV, inward-going 'tail' currents were detectable on returning to the holding potential. Both the nonlinear properties of the steady-state I - V curve and the tail currents were eliminated in solutions nominally free of Ca^{2+} (not shown) or by application of either Cd^{2+} (Fig. 1d, e) or other Ca^{2+} conductance antagonists, including manganese, Co,

nickel and D600, suggesting that Ca^{2+} entry is required for activation of the conductance.

The time-dependent currents generated during the depolarizing commands ('on response') and subsequent tail currents ('off response') probably reflect activation of a single ion conductance mechanism, because analysis of responses to voltage-clamp double-step protocols reveals that the I - V relations and reversal potentials (E_{rev}) for both 'on' and 'off' current responses are the same. E_{rev} cannot be measured directly in the recording conditions normally used (Cl^- -loaded cells) because E_{rev} lies within the potential range for activation of the conductance. Extrapolated 'instantaneous' I - V relations of pairs of I (steady-state and tail currents) and V measurements taken from the data in Fig. 1d, however, give a reversal potential of -17.6 mV (s.e.m. = 0.8 mV, $n = 7$ pairs) and the result for a constant activation step (not shown) was -19.6 mV (s.e.m. = 0.2 mV, $n = 7$ pairs). The steady-state I - V relationship is linear in the -20 to 0 mV range and the amplitudes of the corresponding tail currents are maximal (Fig. 1e), suggesting that the conductance is fully activated in this range. The intercept of the I - V relationship over this range with either the passive I - V relations (extrapolated from data in the -79 mV to -55 mV range) or the I - V relationship in the presence of blocking concentrations of Cd^{2+} is -17 mV (Fig. 1e). This is consistent with our estimate of E_{rev} above. When maximally activated from a holding potential of -55 mV, the magnitude of the Ca -dependent conductance evoked in Fig. 1d was 150 nS (calculated from tail currents and assuming E_{rev} is -17 mV). This Ca -dependent conductance did not inactivate appreciably during depolarizing steps to between -20 and -10 mV for up to 10 s, as evidenced by the constancy of tail current amplitudes following depolarizing commands maintained for this period of time (not shown). Activation at -10 mV could be described by an exponential function with a single time constant of 70 ms (from tail current measurements). The decay of tail currents following repolarization of the membrane from the activation potential reflects deactivation of the

Fig. 2 Time-variant currents remaining in cells bathed in TTX and TEA and after Cs injection are sensitive to E_{Cl} . **a**, Left panel, upper trace: a typical double-voltage-step protocol used to derive estimates of the reversal potential (E_{rev}) of tail currents evoked by a constant activation (depolarizing) voltage step. A series of I responses to V steps is superimposed. Middle trace: superimposed current responses, in normal bathing medium containing 158 mM Cl^- . Lower trace: current responses obtained when Na isethionate was pressure-applied in the vicinity of the cell to reduce $[Cl^-]_o$ from 158 mM to about 37 mM. Right panel: two current records in response to the same voltage command are superimposed to emphasize the difference between the responses recorded in control (trace 1) and low $[Cl^-]_o$ (trace 2). When $[Cl^-]_o$ is reduced, current developed during the depolarizing step is reversed in direction, the tail current increases in amplitude and its decay slows. Current measurements were made as indicated by the arrows. Time-invariant or leak current is represented by triangles and instantaneous currents for the activated conductance state are represented by the open and closed circles. Instantaneous $I-V$ relations were extrapolated to their intercept with that extrapolated from leakage or passive conductances in **b**. Use of such pairs of currents to make this estimate was justified by the identity of the conductance underlying the two components (I_{ss} and I_{tail}) and also necessitated by the lability inherent in both activation and driving force parameters. For example, the steady-state current during depolarization ($\circ$) was not necessarily constant, particularly in low $[Cl^-]_o$ (**a**, left panel, lower trace). $I-V$ plots of the type shown in **b** revealed a progressive collapse of the driving force in this cell when bathed in low $[Cl^-]$ but with no apparent change in the degree of activation (not shown). **b**, Leak $I-V$ plots (currents measured before depolarizing voltage step: triangles) and chord $I-V$ relationships (circles) following activation of the conductance with a depolarizing voltage step. Currents were sampled as indicated in **a**. These data were obtained from another cell for which the driving force was more stable and for which E_{rev} for GABA-activated currents was also calculated to monitor changes in the Cl^- gradient across the membrane. Left panel: E_{rev} in normal medium ($[Cl^-]_o = 158$ mM) is estimated to be +10.8 mV (s.e.m. = 1.4 mV, $n = 6$ $I-V$ pairs). Right panel: ambient NaCl was displaced by close application of Na isethionate in the vicinity of the recorded cell to reduce $[Cl^-]_o$, theoretically to a minimum of 37 mM. E_{rev} now extrapolates to +63.7 mV (s.e.m. = 3.2 mV, $n = 6$ $I-V$ pairs), a change which somewhat exceeds that predicted from the Nernst relationship (that is, +37 mV), for a conductance involving solely Cl^- and assuming that the effective $[Cl^-]_o$ at the membrane is 37 mM. The shift in E_{rev} for GABA-activated current in this cell was about +40 mV (not shown). **c**, Tail currents invert at about -60 mV when recording with Cs_2SO_4 -filled electrodes. Left: activation of a slowly developing outward current by a command voltage step to -20 mV. Similar current responses in other cells were abolished in 0.2 mM Cd^{2+} . Right: tail currents measured, in the same cell, at different membrane potentials (indicated) following an activating (depolarizing) command voltage step. The polarity of the tail current reverses at about -60 mV.



conductance, since instantaneous $I-V$ plots from varying times during tails indicate a progressive reduction in the membrane conductance rather than a change in the driving force for the current. Deactivation is apparently voltage-sensitive, being faster at more hyperpolarized potentials (Fig. 2a) and slower during prolonged or intense activation or with increased $[Ca^{2+}]_o$. This is consistent with the regulation of conductance by Ca^{2+} entry and saturation of Ca-sequestering mechanisms during prolonged or strong depolarizing commands.

This conductance mechanism superficially resembles Ca-dependent K^+ conductances recorded in a variety of excitable membranes in its apparent voltage-sensitivity, time course of activation and relative lack of inactivation^{12,13}. The reversal potential for the current response is not consistent with a K^+ -selective mechanism, however, nor is it consistent with a pure Ca^{2+} current, which would be expected to reverse at quite positive potentials¹⁴. E_{rev} is in fact identical to that of γ -aminobutyric acid (GABA)-activated Cl^- currents recorded in the same cells¹⁵. Ca-dependent 'nonspecific' cation currents, described for other preparations¹⁶⁻¹⁸, also reverse in this potential region (about 0 mV).

We used a double-voltage-step protocol in different ionic conditions to estimate E_{rev} for tail currents (Fig. 2a). Neither complete replacement of extracellular Na^+ with choline, nor raising $[K^+]_o$ from 2.5 mM to 25 mM, nor raising $[Ca^{2+}]_o$ from

4 mM to 20 mM altered E_{rev} measured in this way. We did find, however, that replacement of much of the extracellular Cl^- by the impermeant anion isethionate¹⁹ altered current responses evoked during depolarizing commands, changed the amplitude and time-course of tail currents (Fig. 2a) and shifted the estimated E_{rev} in a depolarizing direction (Fig. 2b). Replacement of Cl^- with isethionate also shifted E_{rev} for GABA-activated currents in a depolarizing direction in the same cell (not shown). Similar sizeable shifts in estimates of E_{rev} have been recorded using methane sulphonate to replace Cl^- . Furthermore, tail currents, measured following activating voltage steps, using SO_4^{2-} -filled electrodes (without loading the cell with Cl^-), reverse in polarity at -60 mV (Fig. 2c), which corresponds well with E_{Cl} in these cells²⁰. E_{rev} averaged -1.7 mV (s.e.m. = 3.5 mV, $n = 20$ from 7 cells) when Cl^- -filled electrodes were used and -44 mV (s.e.m. = 2.3 mV, $n = 9$ from 8 cells) when SO_4^{2-} -filled electrodes were used. These findings taken together are consistent with the hypothesis that the conductance involves primarily, if not exclusively, Cl^- .

Ca-dependent Cl^- conductance mechanisms have been reported in toad oocyte²¹ and salamander photoreceptor membranes²² but their functional roles in these cells are unclear. Ca-dependent depolarizing shifts have also been observed in vasopressinergic neurones cultured from the mammalian hypothalamus¹. In our experiments, it was necessary to eliminate

some of the K^+ conductances in order to reveal the Ca-dependent Cl^- conductance. Activation of K^+ conductances attenuates spread of depolarizing current injections, indicating that the Cl^- conductance may be electrotonically distant from the current-passing microelectrode. Presumably, membrane depolarization increases intracellular Ca^{2+} concentration, $[Ca^{2+}]_i$, by activating Ca^{2+} conductances in plasma (and/or cisternal) membranes. It is well established that Ca^{2+} conductances are present at terminals in the vertebrate central nervous system, where they participate in transmitter release. If the Ca-dependent Cl^- conductance is also present at nerve endings, and provided that E_{Cl} is similar at terminals and cell bodies⁵⁰, then activation of the conductance at this level, by depolarization and/or Ca^{2+} entry, may serve to repolarize the terminal region, thus limiting transmitter release.

We thank Joann Mazzetta for culturing spinal cord neurones, Drs Amy MacDermott and David Lange for helpful criticism and Mary Ann Bragg for typing the manuscript.

Received 7 February; accepted 3 August 1984.

1. Legendre, P., Cooke, I. M. & Vincent J.-D. *J. Neurophysiol.* **48**, 1121-1141 (1982).
2. Dichter, M. A. & Fischbach, G. D. *J. Physiol., Lond.* **267**, 281-298 (1977).
3. Fukada, J., Fischbach, G. D. & Smith, T. G. *Jr Deol Biol.* **49**, 412-424 (1976).
4. MacDonald, J. F. & Schneiderman, H. H. *Brain Res.* **296**, 350-355 (1984).
5. Barrett, E. F. & Barrett, J. N. *J. Physiol., Lond.* **255**, 737-774 (1976).
6. Aldrich, R. W., Getting, P. A. & Thompson, S. H. *J. Physiol., Lond.* **291**, 531-544 (1979).
7. McAfee, D. A. & Tarowsky, P. J. *J. Physiol., Lond.* **290**, 507-523 (1979).
8. Barrett, E. F., Barrett, J. N. & Crill, W. E. *J. Physiol., Lond.* **304**, 251-276 (1980).
9. Ransom, B. R., Neale, E., Henkart, M., Bullock, P. N. & Nelson, P. G. *J. Neurophysiol.* **40**, 1132-1150 (1981).
10. Adams, P. R. in *Advances in Physiological Sciences* Vol. 4 (ed. Salankin, J.) 135-138 (Pergamon, New York, 1981).
11. Puil, E. & Werman, R. *Can. J. Physiol. Pharmacol.* **59**, 1280-1284 (1981).
12. Brown, D. A., Griffith, W. H. & Halliwell, J. V. *J. Physiol., Lond.* **324**, 63P (1982).
13. Segal, M. & Barker, J. L. *J. Neurophysiol.* (in the press).
14. Reuter, H. & Scholz, H. *J. Physiol., Lond.* **262**, 17-47 (1977).
15. Barker, J. L., McBurney, R. N. & MacDonald, J. F. *J. Physiol., Lond.* **322**, 365-387 (1982).
16. Lansman, J. B. *Biophys. J.* **41**, 61a (1983).
17. Colquhoun, D., Neher, E., Reuter, H. & Stevens, C. F. *Nature* **294**, 752-754 (1981).
18. Yellen, G. *Nature* **296**, 357-359 (1982).
19. Chesnoy-Marchais, D. *J. Physiol., Lond.* **342**, 277-308 (1983).
20. Barker, J. L. & Ransom, B. R. *J. Physiol., Lond.* **280**, 331-354 (1978).
21. Barish, M. E. *J. Physiol., Lond.* **342**, 309-325 (1983).
22. Bader, C. R., Bertrand, D. & Schwartz, E. A. *J. Physiol., Lond.* **331**, 253-284 (1982).

The agonist effect of dihydropyridines on Ca channels

A. M. Brown, D. L. Kunze & A. Yatani

Department of Physiology and Biophysics, University of Texas Medical Branch, Galveston, Texas 77550, USA

Dihydropyridines (DHP) have great potential for clinical use because of their inotropic and vasomotor effects. The mechanism of action is unknown although Ca currents have been implicated^{1,2}. Here we report measurements of single channel and whole cell cardiac Ca currents after application of two DHP agonists BAY K 8644 and CGP 28 392. Whole cell Ca currents from individual myocytes were increased and the 50% effective doses (ED_{50}) were similar to those reported for contractility in rabbit aorta and guinea pig heart¹ and catecholamine release from cat adrenal glands³. The measured ED_{50} was also consistent with the apparent dissociation constant (K_d) of a high affinity binding site present in cardiac sarcolemmal vesicles⁴⁻⁷. We propose that the molecular basis for these results is an increase in the probability that a single Ca channel, having opened and closed, will subsequently re-open during membrane depolarization. At high concentrations of BAY K 8644 and in the presence of 96 mM Ba, different effects are observed, primarily a marked prolongation of open time⁸.

Neonatal rat and adult guinea pig ventricular cells were prepared using the methods of refs 9 and 10 respectively. The guinea pigs (300-500 g in weight) were anaesthetized with pentobarbital (50 mg per kg body weight) and the ascending aorta

was cannulated *in situ* under artificial respiration before removing the heart¹¹. The enzyme solution used for dispersion was washed out with 'power soup'¹².

Whole cell currents were recorded using the blown-out patch clamp method¹³ and single channel Ca currents were recorded using the patch gigaseal method in the cell-attached configuration¹³. Cardiac action potentials were recorded in five experiments with standard intracellular microelectrodes. For whole cell Ca currents, the extracellular fluid (in mM) was: NaCl 135, KCl 5.4, $CaCl_2$ 1.8, $MgCl_2$ 1.0, glucose 10, HEPES 5.0 (pH was adjusted to 7.4 with NaOH). The intracellular pipette contained (in mM), KOH 110, aspartic acid 110, KCl 20, HEPES 5, EGTA 0.2, ATP 2 and the pH was adjusted to 7.3 with KOH. The Ca patch pipette contained (in mM): $CaCl_2$ 40, tetraethylammonium chloride (TEACl) 20, KCl 5.5, glucose 20, 4-aminopyridine (4-AP) 5.0, Tris-HCl 55. The pH was adjusted to 7.4 with KOH. The Ba patch pipette contained (in mM): $BaCl_2$ 96, HEPES 10 (pH 7.4). All experiments were done at room temperature (20-22 °C).

Both agonists at low concentrations produced dose-dependent increases in the duration of the ventricular action potential with no changes in the resting potential (Fig. 1D). At an action potential of 0 mV, the duration was increased by as much as 25% whereas at half-peak amplitude the duration was increased by a factor of two or more. The effects occurred within 1 min and were usually complete within 5 min (Fig. 1E). The effects of the agonists were reversible after 15-20 min of washing. The effect of BAY K 8644 (10^{-7} M) on whole cell peak Ca currents was an increase between -20 and +30 mV (Fig. 2A) and similar values were obtained when Cs was substituted for K in the

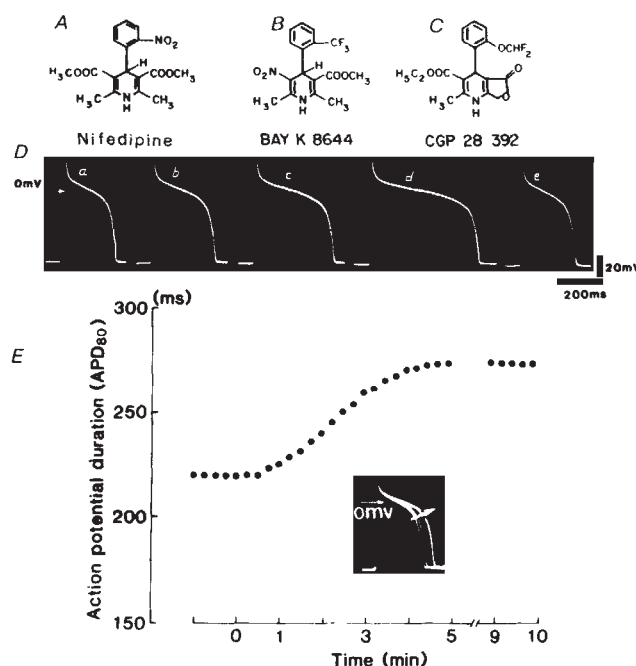
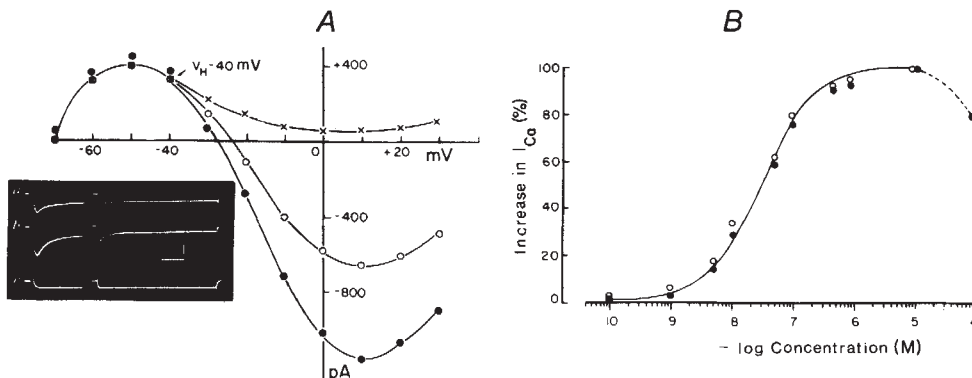


Fig. 1 A-C, Chemical structures of nifedipine, BAY K 8644 and CGP 28 392. The difference between dihydropyridines with antagonist effects and those with agonist effects is that the former have two freely rotating ester groups whereas the latter have only one. The difference between BAY and CGP is that a nitro group replaces an ester group in BAY and one ester group is locked into a rigid ring structure in CGP. D, E, The effects of BAY on the action potential (AP) of single isolated adult guinea pig ventricular cells. D, Dose-dependent changes; a, control; b, 5×10^{-9} M; c, 5×10^{-8} M; d, 5×10^{-7} M and e, recovery 15 min after washing. The major effect is a prolongation of AP duration. E, The time course of the changes in AP measured at 80% repolarization (APD₈₀) produced by 5×10^{-9} M BAY. The inset shows the AP before and 1, 3, 5 and 10 min after adding the drug. Stimulation occurred every 15 s.

Fig. 2 **A**, Effects of BAY K 8644 (10^{-7} M) on Ca current and peak $I-V$ curve of single adult guinea pig ventricular myocytes. $\circ$, Control; $\bullet$, plus BAY K 8644 (10^{-7} M); $\times$, plus Co. The holding potential (V_H) was -40 mV and steps were applied every 10 s. The inset shows the currents in response to steps from -40 to $+10$ mV before and after drug (*a*, *b*) and after blocking calcium current with 3 mM Co (*c*). There are two time bases, left, 10 ms; right, 100 ms. Current calibration was 500 pA. The outward holding current is produced by using a holding potential which is less negative than the zero current potential; it is carried by the inward rectifier. Ca currents for the dose-response curves were measured as the difference in inward current before and after addition of Co. **B**, Dose-response curves of maximum Ca current in adult guinea pig for the CGP agonist ($\circ$) and the BAY agonist ($\bullet$). Each point is the mean value of seven cells for BAY and five cells for CGP. S.E. values were $< \pm 10\%$ for both agents. The solid line extends only to the peak I_{Ca} and is the theoretical curve for one to one agonist-receptor interactions. ED_{50} is 3×10^{-8} M. Similar results were obtained in neonatal rat myocytes. The maximum increase produced by BAY was three- to fourfold whereas that produced by CGP was twofold.



intracellular pipette. At doses of 10^{-7} M or greater, the effects for BAY K 8644 but not CGP 28 392 appeared to be larger at negative test potentials. Inactivation from the peak was quicker, the steady current was larger and tail currents were prolonged. As results of this kind have been attributed to loss of potential control¹⁴, we investigated these effects using neonatal myocytes, which have smaller currents and simpler geometry due to the absence of T-tubules¹⁵. The bath solution (in mM) was: CaCl_2 5, TEACl 20, KCl 5.5, 4-AP 5, Tris-HCl 90, glucose 20; pH was adjusted to 7.3 with KOH. The pipette solution (in mM) was: CsOH 110, aspartic acid 110, CsCl 20, HEPES 5, EGTA 1, ATP 2; pH was adjusted to 7.3 with CsOH. The currents were controlled in these cells and in these conditions, a voltage-dependence of drug action was not observed. Nevertheless we cannot exclude a voltage-dependent effect in adult cells. The effects of alteration in the state of the channel were investigated in inactivated and rested channels. To study steady-state inactivation, holding potentials as negative as -70 mV were required. This could reactivate Na currents which were abolished at the usual holding potential of -40 mV. To avoid this, tetrodotoxin (TTX) was added in doses of 5×10^{-5} M to completely block Na currents¹⁶. Steady-state inactivation measured using prepulses of several seconds duration combined with test pulses to either $+10$ or -10 mV was unaffected by agonists. We did not observe the blocking action reported for syncytial cardiac Purkinje preparations¹⁷ at the ED_{50} using holding or conditioning potentials of -40 mV.

The effects were specific for Ca currents. The inward rectifier current activated by hyperpolarizing potentials to -100 mV and the outward currents activated at potentials above $+20$ mV were unaffected. The β -blocker propranolol at a concentration of 5×10^{-6} M had no effect; thus β -receptors which when activated by noradrenaline also increase Ca currents in these cells^{18,19} were not involved. The Ca currents were eliminated by the specific DHP Ca antagonist nitrendipine in doses of 10^{-5} M or by substitution of Co for Ca extracellularly. The dose-response curves for both agonists were similar (Fig. 2B). At lower doses the effects were reversible; at higher doses recovery was partial. This may be partially due to the fact that Ca currents in small cells run down with time^{20,21}. A maximum effect at an agonist concentration of 10^{-5} M was also observed when contraction of aortic smooth muscle or heart¹ or noradrenaline release³ was used as the observable response.

To examine the mechanism of action at a molecular level, we studied single Ca channels. The trials consisted of up to 500 identical voltage steps 50–500 ms duration delivered at intervals of 4–12 s. Long periods of recording during control were associated with a progressive loss of activity and it was useful to construct a trace by trace 'diary' of the single channel activity. The diary consisted of measurements of the percentage of open

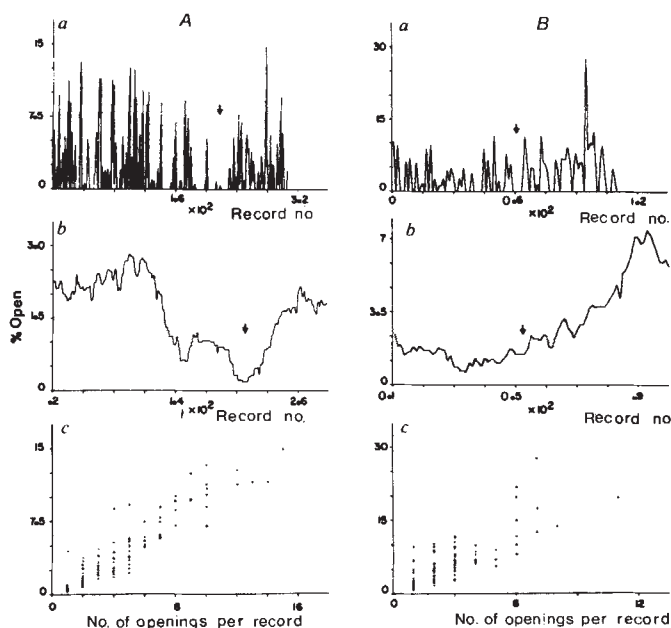
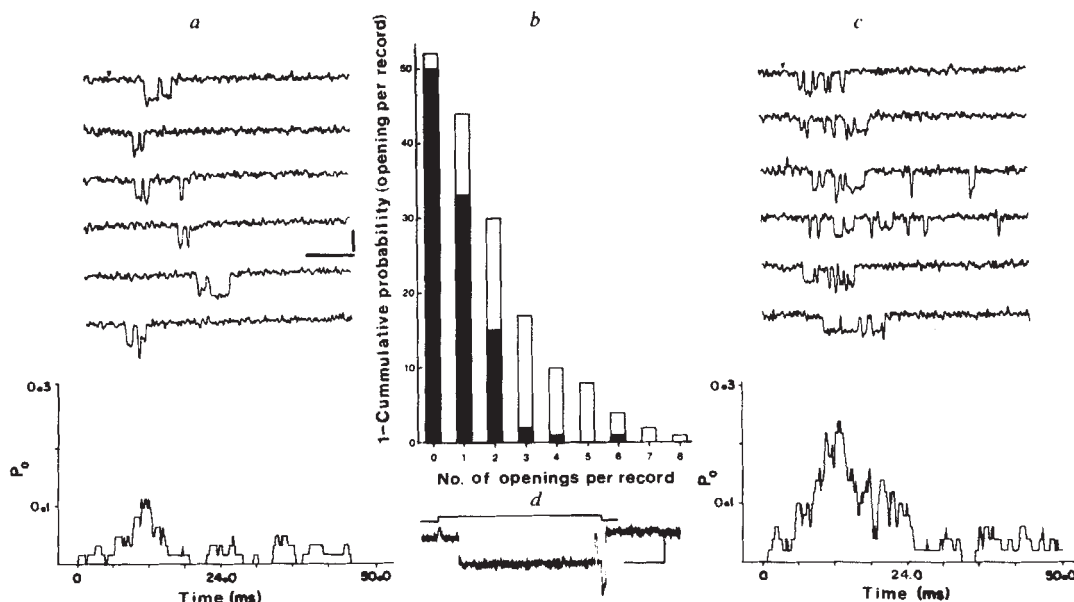


Fig. 3 Single channel activity from two cell-attached patches on neonatal rat ventricular myocytes during control and after addition (marked by arrows) of BAY K 8644 (**A**) and CGP 28 392 (**B**) in doses of 5×10^{-7} M. The patch pipette contained 40 mM Ca. In **B**, the potential step was applied at 0.25 Hz for 70 ms. Resting potential, V_R , estimated (from whole cell measurements) was -50 mV. The patch potential was -20 mV, and the step potential (V_{step}) was $+50$ mV. In **A** the step was 100 ms in duration, holding potential was -20 mV, and V_{step} was 40 mV; repetition rate was 0.2 Hz. Records were low pass filtered at 1 kHz using a zero phase, nonringing, digital filter; sampling rate was 50 μ s per point. **a**, Percentage of open time against the record number. **b**, Smoothed versions of **a**, where smoothing is done by averaging at each point the surrounding 41 and 21 points in **A** and **B** respectively. Note that in **B**, rundown during control was small whereas in **A** it was substantial. **c**, The relationship between the percentage of open time (ordinate) and number of openings per record (abscissa). The data were fitted by linear regressions with $r = 0.84$ and 0.69 , respectively. The slopes correspond to open times per opening of 0.88 and 1.0 ms in **A** and **B** respectively. Large deviations due to prolonged openings are not present.

time per record and the number of openings per record (Fig. 3). Figure 3 shows that in control experiments the activity was cyclic; it also confirms the occurrence of rundown. Addition of agonist was always associated with a large increase in activity although the cyclical pattern persisted. Similar results were observed in three additional experiments. At an agonist

Fig. 4 Effects of agonist CGP 28 392 on single channel Ca currents (*a*, control; *c*, after drug). Control traces are representative of record numbers 25–65 and traces after drug are representative of record numbers 80–120 shown in Fig. 3*B*. Arrows indicate onset of potential step. There were two channels in this patch (sixth trace in *a*, third trace in *c*). Bars are 0.4 pA and 15 ms. Linear components of capacitance and leakage have been subtracted. *c* was taken 3 min after addition of 5×10^{-7} M CGP 28 392 to the bath. Below *a* and *c* are the averaged currents for the record numbers referred to above— P_0 is the opening probability. *b*, Agonist effect on the distributions of openings per trace. ■, Control; □, plus CGP 28 392 (5×10^{-7} M). The ordinate is 1—the cumulative discrete probability density function of the number of openings per record. *d*, Effect on a single channel of 10^{-5} M BAY. Pipette contained isotonic Ba. V_H was the resting potential of -50 mV, V_{step} $+40$ mV, pulse duration 100 ms. Calibration, 1 pA and 20 ms. Slope conductance calculated from the closure following return of the voltage step was 30 pS, which is about 20% greater than the usual value in isotonic Ba (ref. 18).



concentration of 5×10^{-7} M with 40 mM Ca in the patch pipette, the relationship between percentage of open time and number of openings per record was monotonic and highly correlated. In 1,536 events there was no evidence of greatly prolonged open times such as those observed in Fig. 4*d*. Unitary amplitudes were unchanged and the major effect was an increase in the average number of openings per sample due to the fact that re-openings became more frequent (Fig. 4). When the individual single channel records were summed and averaged, they increased and decreased in the same way as the whole cell currents, and the increase due to agonist was consistent with the increase observed in whole cell current recordings (Fig. 4). The open time distributions were fitted by a single exponential function and were unaffected by agonist. The mean value of open time was 1.36 ± 0.76 ms ($n=815$) for the control experiments and 1.47 ± 0.98 ms ($n=721$) for agonist when holding potential (V_H) = resting potential (V_R) -20 to -40 mV and the voltage step (V_{step}) was $+30$ to $+60$ mV.

The mean values for waiting times (latencies to first opening) were unchanged. The openings occurred in bursts as already reported^{20,22–24}, and these gave rise to short and long closed times. The burst duration and gaps within bursts were unaffected and the mean closed times were also unchanged. Although increased re-opening should reduce the mean closed times, for pulsed data, the last closed times were not determined. More of the longer closed times are discarded in the control experiment so that the drug effect is obscured. When high agonist concentrations (10^{-5} M) were used with 96 mM Ba in the patch pipette, repetitive activity was markedly intensified and greatly prolonged openings appeared⁸ (Fig. 4*d*). These events gave a bimodal distribution to the plots of percentage open time versus number of openings per record.

Bath applications of nitrendipine in doses which blocked the whole cell current responses also blocked the single channel responses. It is not surprising that DHPs gain access to the membrane patch via extra-patch membrane as these agents are highly lipophilic. However, it is not known whether equilibrium conditions for drug distribution are attained in the patch.

The effects on single channels were due to an increase in the probability of re-opening. We cannot exclude an increase in the number of activatable Ca channels, but the results suggest that

much of the increase in whole cell current is attributable to the increased probability of re-opening. Two possibilities are that progression to an absorbing inactivated state was inhibited or that re-opening from a closed state was enhanced. The absence of an effect on steady-state inactivation might favour the latter interpretation. The concentrations we used indicate that the effects were mediated by the high-affinity binding sites. In these circumstances the agonists did not change channel structure sufficiently to increase unit conductance or duration in the open state. The dramatic changes observed at 10^{-5} M concentrations of agonist⁸ indicate a considerable change in channel structure and might be mediated by the low-affinity binding sites described for these agents^{4,6}. They might also be associated with the falling phase of the dose-response curves.

We thank Dr A. Scriabine (Miles Laboratories) and Dr G. Weiss (Ciba-Geigy) for supplies of the agonists. This work was supported in part by NIH grants NS 11453 and HL 25145.

Received 24 April; accepted 25 July 1984.

- Schramm, M., Thomas, G., Towart, R. & Franckowiak, G. *Nature* **303**, 535–537 (1983).
- Glossman, H., Ferry, D. R., Lübbecke, F., Mewes, R. & Hofmann, F. *Trends pharmac. Sci.* **3**, 431–437 (1982).
- Garcia, A. G. *et al.* *Nature* **309**, 69–71 (1984).
- Janis, R. A., Rampe, D., Sarmiento, J. G. & Triggle, D. J. *Biochem. biophys. Res. Commun.* (in the press).
- Sarmiento, J. G., Janis, R. A., Rampe, D. & Triggle, D. J. *Fedn Proc.* **43**, 448 (1984).
- Vaghy, P. L., Grupp, I. L., Grupp, G. & Schwartz, A. *Circulation Res.* (in the press).
- Grupp, I. L. *et al.* *Fedn Proc.* **43**, 937 (1984).
- Hess, P., Lansman, J. B. & Tsien, R. W. *Biophys. J.* **45**, 394a (1984).
- Mark, G. E. & Strasser, F. F. *Expl Cell Res.* **44**, 217–233 (1966).
- Powell, T., Terrar, D. A. & Twist, V. W. *J. Physiol., Lond.* **302**, 131–153 (1980).
- Taniguchi, J., Kokubun, S., Noma, A. & Irisawa, H. *Jap. J. Physiol.* **31**, 547–558 (1981).
- Isenberg, G. & Klockner, U. *Pflügers Arch. ges. Physiol.* **395**, 6–18 (1982).
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85–100 (1981).
- Johnson, E. A. & Lieberman, M. A. *Rev. Physiol.* **33**, 479–532 (1971).
- Moses, R. L. & Kasten, F. H. *Cell Tissue Res.* **203**, 173–180 (1979).
- Josephson, I., Sanchez-Chapula, J. & Brown, A. M. *Circulation Res.* **54**, 157–162 (1984).
- Sanguinetti, M. C. & Kass, R. S. *Biophys. J.* **45**, 394a (1984).
- Reuter, H., Stevens, C. F., Tsien, R. W. & Yellen, G. *Nature* **297**, 501–504 (1983).
- Cachelin, A. B., de Peyer, J. E., Kokubun, S. & Reuter, H. *Nature* **304**, 462–464 (1983).
- Fenwick, E. M., Marty, A. & Neher, E. *J. Physiol., Lond.* **331**, 599–635 (1982).
- Hagiwara, S. & Ohmori, H. *J. Physiol., Lond.* **336**, 649–661 (1983).
- Lux, H. D. & Nagy, K. *Pflügers Arch. ges. Physiol.* **391**, 252–254 (1981).
- Brown, A. M., Camerer, H., Kunze, D. L. & Lux, H. D. *Nature* **299**, 156–158 (1982).
- Cavalié, A., Ochi, R., Pelzer, D. & Trautwein, W. *Pflügers Arch. ges. Physiol.* **398**, 284–297 (1983).

Immunochemical tests of acetylcholine receptor subunit models

Jon Lindstrom*, Manuel Criado*,
Susan Hochschwender*, J. Lawrence Fox†
& Virender Sarin†

* The Salk Institute for Biological Studies, PO Box 85800, San Diego, California 92138, USA

† Department of Molecular Biology, Abbott Laboratories, Abbott Park, North Chicago, Illinois 60064, USA

Acetylcholine receptors of fish electric organs and mammalian skeletal muscle comprise four structurally homologous glycoprotein subunits in the mole ratio $\alpha_2\beta\gamma\delta$ (refs 1–4). All four subunits have leader sequences and are exposed on both sides of the membrane^{5,6}. From amino acid sequencing, three groups^{5,7,8} have predicted that each subunit has four hydrophobic α -helical transmembranous domains. Because the N-terminus of each subunit is thought to remain on the extracellular surface after cleavage of the leader sequence, this model predicts that the N- and C-termini are both on the extracellular side. An alternative model proposed by two other groups^{9,10} predicts that there is, in addition, a fifth amphipathic transmembranous domain which would place the C-terminus on the cytoplasmic side. Here, using anti-subunit sera and monoclonal antibodies and their reaction with synthetic subunit peptides, we demonstrate that the C-terminus is in fact on the cytoplasmic surface. We also show that, contrary to other predictions¹¹, the most hydrophilic sequence on the extracellular domain of α -subunits is not the main immunogenic region.

More than half of the antibodies made against native acetylcholine receptors (AChR) are directed at a small part of the extracellular surface of the α -subunits called the main immunogenic region (MIR)^{12–14}. It has been reported that the most hydrophilic part of an amino acid sequence is usually a prominent antigenic determinant¹⁵. Therefore, it has been suggested that the most hydrophilic sequence on the predicted extracellular domain of α -subunits would form the main immunogenic region, although we show here that this is not correct.

Peptides were synthesized corresponding to the N- and C-terminal 10 amino acids of each subunit (except for δ , where the C-terminal peptide extended from C-3 to C-12 because of synthetic difficulties with the penultimate cysteine) and to the predicted main immunogenic region on the α -subunit (residues 159–169; ref. 11) (Bachem). At the unnatural terminus of each peptide, a tyrosine was added so that the peptides could be labelled with ¹²⁵I.

Anti-receptor monoclonal antibodies from our library (refs 12–14 and unpublished data) and antisera to native receptor from *Torpedo californica* and its purified denatured subunits were tested for reaction with these ¹²⁵I-labelled synthetic peptides (see Table 1). The results indicate that the N-termini of the chains are negligible immunogens in the intact receptor. Antisera to SDS-denatured receptor subunits also showed negligible reactions with the N-termini, however, the C-termini of the β -, γ - and, to some extent, δ -subunits were significantly immunogenic. Comparison of the titres of these anti-subunit sera to intact receptor (all $\sim 2.5 \mu\text{M}$) with their titres against C-terminal peptides suggests that $\sim 20\%$ of antibodies to β -subunits, $\sim 6\%$ of antibodies to γ -subunits, and $\sim 0.4\%$ of the antibodies to δ -subunits are directed at the C-terminus. Titres against intact receptor were measured by indirect immune precipitation in units of mol α -bungarotoxin binding sites bound per litre of solution, and converted to approximate molar concentration of antibody by assuming that all of the receptors were present as dimers (containing four binding sites) and dividing the titre by four. Titres against peptides were measured by indirect immune precipitation in units of mol peptide bound per litre of solution, and converted to approximate molar concentration of antibody by assuming that peptides were bound to both of the antibody binding sites and dividing the titre by two. The fraction of antibodies to the C-terminus was calculated by dividing the concentration of antibodies to the peptide by the concentration of antibodies to the intact receptor. Three monoclonal antibodies to β -subunits were found which recognized the C-terminus. These monoclonal antibodies were initially selected by S. Tzartos *et al.* and were used in refs 16–18. They are the result of the fusion of spleen cells from rats immunized with β -subunits with the mouse myeloma S194 as previously described^{12–14}. Their subclasses are 163-IgM, 170-IgG2a and 172-IgG1.

Antibodies which bound to C-termini were tested to determine whether they bound to the extracellular surface or the cytoplasmic surface of the receptor. Receptor-rich membrane vesicles from *Torpedo* electric organ are sealed rightside-out¹⁹, but can be permeabilized using saponin or lithium diiodosalicylate (LIS)²⁰. The ability of native or permeabilized vesicles to adsorb antibodies to the C-termini was tested (Fig. 1, Table 2). These results clearly demonstrate that the C-termini of β -, γ - and δ -subunits are on the cytoplasmic surface of the receptor and are accessible to antibody binding only when receptor-rich membrane vesicles have been permeabilized to permit entry of the antibody molecules. Pretreatment of the vesicles with saponin (0.3% or 3%) also permitted the vesicles to adsorb antibodies to the C-termini (Fig. 1). Previously, we found 11 monoclonal antibodies to receptor which required saponin to bind to receptors in the postsynaptic membrane of muscle and then showed

Table 1 Reaction of antibodies with synthetic peptides from receptor subunits

	[(Antibody c.p.m. divided by blank c.p.m.) - 1]								
	α_{1-10}	$\alpha_{159-169}$	$\alpha_{428-437}$	β_{1-10}	$\beta_{460-469}$	γ_{1-10}	$\gamma_{480-489}$	δ_{1-10}	$\delta_{490-499}$
Anti-AChR serum	0	0	0	0	0	0	7	0	0
Anti- α serum	0	3	0	0	0	0	0	0	0
Anti- β serum	0	0	0	0	44	0	0	0	0
Anti- γ serum	0	0	0	0	3	0	147	0	1
Anti- δ serum	0	0	0	0	0	0	30	0	2
18 Anti- α antibodies	0	0	0						
6 Anti-MIR antibodies		0							
40 Anti- β antibodies				0	0				
mAb 163				0	78				
mAb 170				0	41				
mAb 172				0	182				
10 Anti- γ antibodies						0	0		
20 Anti- δ antibodies								0	0

Peptides were labelled with ¹²⁵I to specific activities of about 3×10^{18} c.p.m. mol⁻¹ by the same basic procedure reported for labelling α -bungarotoxin³⁰. After dilution to 1×10^{-9} M (in 0.5% Triton X-100, 100 mM NaCl, 10 mM NaN₃, 10 mM Na phosphate buffer pH 7.5), triplicate 1-ml aliquots were incubated overnight with 5 μ l of antisera or 2 μ l of monoclonal antibody plus 3 μ l of normal rat serum. After precipitation with goat anti-rat IgG, the washed pellets were counted. Normal rat serum was used as a blank. MIR, main immunogenic region.

Table 2 Inhibition of antibody binding to C-terminal peptides by receptor in various forms

	% Inhibition of binding by 3×10^{-8} M AChR		
	Native vesicles	LIS-permeabilized vesicles	Soluble receptor
Anti- β serum	0	68	71
Anti- β mAb 163 (IgM)	3	11	31
Anti- β mAb 170 (IgG)	0	74	55
Anti- β mAb 172 (IgG)	0	52	59
Anti- γ serum	0	39	45
Anti- δ serum	0	44	85

Inhibition experiments were conducted as described for Fig. 1.

by electron microscopy using colloidal gold-labelled anti-antibody that in every case the antibodies bound to the cytoplasmic surface of the membrane¹⁶. Saponin treatment was less effective than LIS treatment (Fig. 1). This observation can be explained²⁰ by the fact that saponin only makes holes in the membrane through which antibodies can enter the vesicle, whereas LIS not only disrupts the vesicle, but also removes a protein (molecular weight 43,000) which is known to associate with the cytoplasmic surface of β -subunits²¹. This observation further supports our proposal that antibodies to the C-termini of the subunits bind to the cytoplasmic surface of the receptor. To test this conclusion further, we prepared antisera against the C-terminal peptides of all four subunits and showed that LIS-extracted but not native receptor-rich vesicles could adsorb those antisera²².

Our results conclusively demonstrate that models^{4,7,8} which place the C-termini of receptor subunits on the extracellular surface of the postsynaptic membrane are incorrect. Those models^{9,10} which propose a fifth amphipathic domain may be correct. R. Stroud and co-workers have reached a similar conclusion using electron microscopy to localize binding of antisera to the C-terminus of δ -subunits (personal communication). A fifth amphipathic transmembranous domain is especially appealing because it provides a hydrophilic surface for lining the cation channel through the receptor, which must have a hydrophilic surface to accommodate the high flow rate of hydrated cations known to occur through the channel²³.

It was proposed that the main immunogenic region of the receptor is formed by the most hydrophilic sequence of the α -subunits (amino acids 161–166)¹⁷. However, Table 1 shows that the synthetic peptide α 159–169 reacts only slightly with antisera to α -subunits and not at all with antisera to native receptor or with a collection of monoclonal antibodies to the

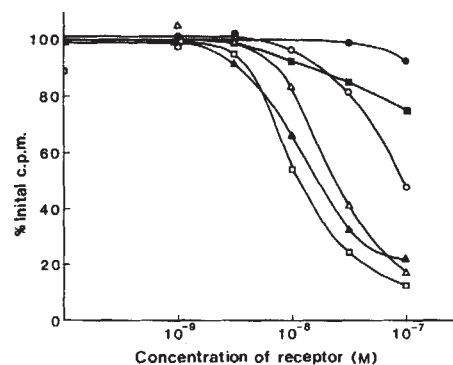


Fig. 1 Inhibition of binding of monoclonal antibody 172 to the ^{125}I -labelled C-terminal peptide of β -subunits. ●, Native vesicles; ■, native vesicles plus saponin; ○, pH 11 extracted vesicles; ▲, soluble receptor; △, pH 11 and Li diiodosalicylate (LIS)-extracted vesicles; □, denatured soluble receptor. Native receptor-rich membrane vesicles¹⁹ were permeabilized with 0.3% saponin¹⁶, or extracted with pH 11²⁹, or extracted first with pH 11 and then with 10 mM LIS²⁰. Affinity-purified solubilized receptor³⁰ was denatured with 1% SDS and then diluted to 0.15% SDS in 0.5% Triton X-100. Triplicate 1-ml aliquots of ^{125}I -labelled C-terminal peptide at 1×10^{-10} M (in 100 mM NaCl, 10 mM NaN_3 , 10 mM Na phosphate pH 7.5) were mixed with the indicated concentrations of receptor (measured as α -bungarotoxin binding sites³⁰). Antibody was added overnight at a concentration of 2×10^{-10} M (measured as α -bungarotoxin binding sites precipitated per litre) in the presence of 5 μl of normal rat serum as carrier, then goat anti-rat IgG was added. The resulting pellet was washed and counted.

main immunogenic region which react with denatured α -subunits. The main immunogenic region is a conformation-sensitive region—antibodies to it react better with native receptor than with SDS-denatured α -subunits. However, as these antibodies react with denatured α -subunits, they should bind to a synthetic peptide corresponding to the main immunogenic region. Therefore, the sequence α 159–169 does not form the main immunogenic region. Juillerat *et al.* recently reached a similar conclusion²⁴.

To test this conclusion further, a larger peptide from the same region of the α -subunit (amino acids 152–167) was synthesized and coupled to keyhole limpet haemocyanin by glutaraldehyde, for use as an immunogen. Spleen cells from an immunized rat were fused with S194 mouse myeloma cells as previously described^{12,13,19} to produce a hybridoma secreting a monoclonal antibody designated mAb 236 (M.C. *et al.*, in preparation). This antibody bound to α -subunits but not to β -, γ - or δ -subunits

Table 3 Competition between monoclonal antibody to α 152–167 and monoclonal antibodies to MIR for binding to AChR in an ELISA

	Additivity index (raw data)			
	mAb 236	mAb 6	mAb 35	mAb 203
mAb 236 (anti- α 152–167)	(0.50 $\pm$ 0.04)	82 (0.80 $\pm$ 0.03)	96 (0.99 $\pm$ 0.08)	97 (1.34 $\pm$ 0.12)
mAb 6 (anti- <i>Torpedo</i> MIR)		(0.38 $\pm$ 0.04)	11 (0.51 $\pm$ 0.05)	–32 (0.42 $\pm$ 0.02)
mAb 35 (anti-eel MIR)			(0.53 $\pm$ 0.04)	–12 (0.61 $\pm$ 0.04)
mAb 203 (anti-human MIR)				(0.86 $\pm$ 0.06)

Receptor was bound to microtitre dishes (Immulon 1, Dynatech) by adding 50 μl per well of a 2×10^{-8} M solution in 0.01 M Na bicarbonate buffer, pH 9.5 overnight at 4 °C. Unreacted sites on the plates were quenched by adding 200 μl of 0.5% bovine serum albumin in 5% Tween 20 for 15 min. This was followed by four washes with the same solution. Monoclonal antibodies at concentrations chosen to just saturate this assay after overnight incubation at 4 °C, were added in the same solution. Monoclonal antibody binding was measured using a monoclonal antibody to rat IgG κ -chains (MAR 18.5) (ref. 26) coupled to glucose oxidase²⁷. The raw data shown are the absorbance in the ELISA for each monoclonal antibody or combination of monoclonal antibodies corrected by subtracting the absorbance of blank wells. The additivity index (AI)²⁸ is calculated as follows: $\text{AI} = 100 ((2A_{1+2}/A_1 + A_2) - 1)$ where A_1 is the absorbance with the first monoclonal antibody alone, A_2 is the absorbance with the second monoclonal antibody alone, and A_{1+2} is the absorbance with the two monoclonal antibodies together. A low value of the additivity index indicates that the two monoclonal antibodies cannot bind simultaneously and compete for binding, whereas a value near 100 indicates that both monoclonal antibodies can bind to receptor simultaneously and do not compete for binding. Combinations of monoclonal antibodies to the main immunogenic region (MIR) give low values of the additivity index. The absorbance in the ELISA of combinations of either mAb 6 or mAb 35 with mAb 203, is close to the absorbance of mAb 6 or mAb 35, probably because mAb 6 was made against receptor from *Torpedo* electric organ¹², and mAb 35 was made against receptor from *Electrophorus* electric organ¹³, whereas mAb 203 was made against receptor from human muscle¹⁴, thus monoclonal antibodies 6 and 35 have higher affinity for the *Torpedo* receptor which was used as the antigen in this assay. Combinations of mAb 6, 35 or 203 with mAb 236 raised to the synthetic peptide α 152–167 give absorbances approximating the sum of the absorbances of the monoclonal antibodies added individually, showing that monoclonal antibodies bound to the MIR do not impair binding of mAb 236 to α 152–167.

bound to enzyme-linked immunosorbent assay (ELISA) wells (data not shown). Similarly, this antibody bound to intact receptor by ELISA and this binding was inhibited by both the synthetic peptides α 159-169 and α 152-167. Using this system, we tested whether mAb 236 competed for binding to the receptor with antibodies to the main immunogenic region. Table 3 shows that monoclonal antibodies to the main immunogenic region compete with each other for binding to receptor, whereas mAb 236 and monoclonal antibodies to the main immunogenic region can bind to receptors simultaneously. This shows that not only is the sequence 161-166 not the main immunogenic region, it is not even immediately contiguous to this region.

The studies described here reveal that predicting protein structure from cDNA sequences (the method used to predict the two

alternative models^{4,7-10} investigated here) is difficult and that immunochemical techniques are a useful approach for testing the predictions of such studies. Further studies with anti-peptide antibodies and other methods will be required to refine models of subunit structure and to account for the conformation changes which have been detected during the maturation and assembly of receptor subunits²⁵.

We thank John Cooper, Doug Schneider and Lorene Langeberg for technical assistance and the NIH (NS11323), the Muscular Dystrophy Association, the McKnight Foundation, the Alexander S. Onassis Public Benefit Foundation, and the Los Angeles and California Chapters of the Myasthenia Gravis Foundation for support. S.H. is supported by a Muscular Dystrophy Association postdoctoral fellowship.

Received 26 March; accepted 8 August 1984.

1. Lindstrom, J., Merlie, J. & Yogeewaran, G. *Biochemistry* **18**, 4465-4470 (1979).
2. Reynolds, J. A. & Karlin, A. *Biochemistry* **17**, 2035-2038 (1978).
3. Raftery, M., Hunkapiller, M., Strader, C. & Hood, L. *Science* **208**, 1454-1457 (1980).
4. Conti-Tronconi, B., Hunkapiller, M., Lindstrom, J. & Raftery, M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6489-6493 (1982).
5. Noda, M. *et al. Nature* **302**, 528-532 (1983).
6. Anderson, D. & Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5598-5602 (1981).
7. Claudio, T., Ballivet, M., Patrick, J. & Heinemann, S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1111-1115 (1983).
8. Devillers-Thiery, A., Giraudat, J., Bentabollet, M. & Changeux, J. P. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2067-2071 (1983).
9. Guy, R. *Biophys. J.* **45**, 249-261 (1983).
10. Finer-Moore, J. & Stroud, R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 155-159 (1984).
11. Noda, M. *et al. Nature* **299**, 793-797 (1982).
12. Tzartos, S. & Lindstrom, J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 755-759 (1980).
13. Tzartos, S., Rand, D., Einarson, B. & Lindstrom, J. *J. biol. Chem.* **256**, 8635-8645 (1981).
14. Tzartos, S., Langeberg, L., Hochschwender, S. & Lindstrom, J. *FEBS Lett.* **158**, 116-118 (1983).

15. Hopp, T. & Woods, K. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3824-3828 (1981).
16. Sargent, P., Hedges, B., Tsavaler, L., Clemmons, L., Tzartos, S. & Lindstrom, J. *J. Cell Biol.* **98**, 609-618 (1983).
17. Swanson, L. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 4532-4536 (1983).
18. Gullick, W. & Lindstrom, J. *Biochemistry* **22**, 3801-3807 (1983).
19. Elliott, J. *et al. Biochem. J.* **185**, 667-678 (1980).
20. Froehner, S., Douville, K., Klink, S. & Culp, W. *J. biol. Chem.* **258**, 7112-7120 (1983).
21. Burden, S., DePalma, R. & Gottesman, G. *Cell* **35**, 687-692 (1983).
22. Ratnam, M. & Lindstrom, J. *Biochem. biophys. Res. Commun.* **122**, 1225-1233 (1984).
23. Lewis, C. & Stevens, C. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6110-6113 (1983).
24. Juillerat, M., Barkas, T. & Tzartos, S. *FEBS Lett.* **168**, 143-148 (1984).
25. Merlie, J. P. & Lindstrom, J. *Cell* **34**, 747-757 (1983).
26. Lanier, L. L., Gutman, G. A., Lewis, D. E., Griswold, S. T. & Warner, N. L. *Hybridoma* **1**, 125-131 (1982).
27. Johnson, R. & Nakamura, R. in *Immunoassays: Clinical Laboratory Techniques for the 1980s* 141-156 (Liss, New York, 1980).
28. Friguet, B., Djavadi-Ohanian, L., Pages, J., Bussard, A. & Goldberg, M. *J. immun. Meth.* **60**, 351-358 (1983).
29. Neubig, R., Krodell, E., Boyd, N. & Cohen, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 690-694 (1979).
30. Lindstrom, J., Einarson, B. & Tzartos, S. *Meth. Enzym. immunochem. Tech.* **74**, 432-460 (1981).

Photoreceptor-specific degeneration caused by tunicamycin

Steven J. Fliesler, Laurence M. Rapp
& Joe G. Hollyfield

Cullen Eye Institute, Baylor College of Medicine, Houston,
Texas 77030, USA

The antibiotic tunicamycin inhibits the biosynthesis of *N*-acetylglucosaminylpyrophosphoryl polyisoprenol¹⁻³, a key intermediate in the formation of the asparagine-linked oligosaccharides of glycoproteins⁴. The effects of tunicamycin have been studied in various biological systems⁵, primarily with the aim of elucidating the role of the carbohydrate moieties in the cellular function of glycoproteins⁶. Rhodopsin, the visual pigment of retinal rod photoreceptor cells, is a membrane glycoprotein which consists of a single polypeptide chain (opsin) to which a chromophoric prosthetic group (11-*cis*-retinaldehyde) and two asparagine-linked oligosaccharide chains are covalently attached⁷. The glycosylation of opsin can be blocked with tunicamycin *in vitro* in conditions where polypeptide synthesis is only slightly decreased^{8,9}. We have reported⁵ that tunicamycin can disrupt the normal assembly of rod outer segment membranes *in vitro* without significantly inhibiting the biosynthesis or intracellular transport of opsin. Here we report that intraocular injection of tunicamycin produces a photoreceptor-specific degeneration characterized by (1) progressive shortening of rod outer segment, (2) decreased electroretinogram amplitudes, (3) cessation of rod outer segment membrane assembly, and (4) eventual photoreceptor cell death.

Frogs (*Rana pipiens*, 20-25 g each) were maintained at 23°C under cyclic lighting (12 h light-12 h dark) in a metabolic incubator. In each animal, one eye was injected with 1.0 µg of tunicamycin (Calbiochem; in 10 µl dimethyl sulphoxide (DMSO)); the contralateral (control) eye was injected with 10 µl DMSO alone. At various times (1, 4, 8 and 20 days), animals were dark-adapted, immobilized, and electroretinogram (ERG) measurements were performed as described previously¹⁰. The

light intensity necessary to obtain a criterion *b*-wave response of 30 µV was defined as the threshold (I_t , expressed in -log units). Following ERG measurements, animals were exposed to normal room illumination for 1 h and the eyes were then removed and processed for light microscopy¹¹.

No differences between the control and tunicamycin-injected eyes were observed with respect to maximal *b*-wave amplitude one day after injections (Fig. 1*a,b*). However, a small increase in threshold was observed in tunicamycin-injected eyes relative to controls: $\Delta I_t = 0.6$ (Fig. 1*a,b*). By day 4, a tunicamycin-induced decrease in maximal *b*-wave amplitude to 73.6% of control (Fig. 1*d*) and increase in threshold, $\Delta I_t = 1.6$ (Fig. 1*c,d*) were clearly evident. Further decreases in the maximal *b*-wave amplitude to 31.6% of control (Fig. 1*f*) and increases in threshold to $\Delta I_t = 3.4$ (Fig. 1*e,f*) were observed by day 8. During the time course of the decrease in maximal *b*-wave amplitude, the ERG *a*-wave amplitude also declined progressively. Finally, no ERG responses could be elicited from eyes 20 days after the tunicamycin injection (Fig. 1*h*), in contrast to control eyes (Fig. 1*g*). Linear regression analyses of ERG amplitudes as a function of post-injection time indicated that the decrease in maximal *a*-wave and *b*-wave amplitudes occurred at a rate of approximately 5% per day (correlation coefficient, 0.98), relative to controls.

The extent of retinal degeneration induced by tunicamycin was assessed by light microscopy. One day after the tunicamycin injection, no gross morphological changes were apparent relative to contralateral controls (Figs 2*a, 3*). However, by 4 days post-injection (Figs 2*b, 3*), rod outer segments in tunicamycin eyes (29.1 ± 3.7 µm) were reduced in length by nearly 23% ($P < 0.05$, $n = 3$) relative to those of controls (37.9 ± 5.5 µm). Furthermore, retinal pigment epithelium cells in the tunicamycin-injected eyes were laden with phagosomal debris (arrows in Fig. 2*b*). Eight days after the tunicamycin injection, photoreceptor degeneration had progressed markedly (Figs 2*c, 3*). Rod outer segments in the tunicamycin-injected eyes (21.8 ± 1.1 µm) were shortened to 48% of control length ($P < 0.001$, $n = 3$). The retinal pigment epithelium cells exhibited several pathological features, including hypertrophy and rounding, clustering of pigment granules in the apical cytoplasm, retraction of apical

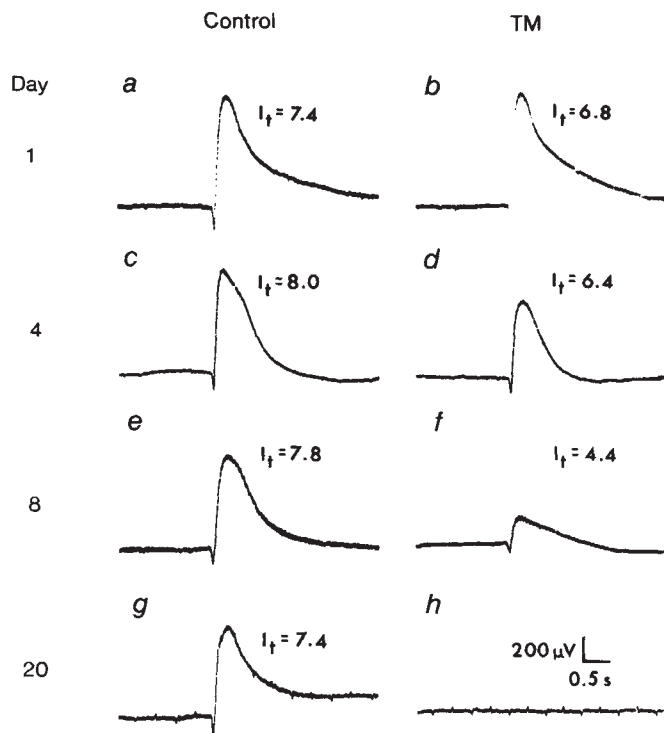


Fig. 1 Electrophoretogram (ERG) recordings obtained following intraocular injection of frogs with DMSO (control: *a, c, e, g*) or tunicamycin (TM: *b, d, f, h*).

Methods: Groups of frogs (3 per time point) were anaesthetized by immersion in 0.5% MS-222 (Sigma) and injected in one eye with a tunicamycin solution (1 μ g, in 10 μ l DMSO) and in the contralateral eye with DMSO (10 μ l). Injections (in the vitreous, at the pars plana) were performed under a dissecting microscope, using a Hamilton syringe fitted with a 32-gauge needle. Dark adaptation, immobilization by injection with curare and isoproteranol, and ERG measurements were performed as described previously¹⁰. Data from the experimental and control eyes are shown for a single animal representative of the group at the given post-injection time. Threshold values (I_t , in $-\log$ units) are given in each panel.

microvilli, and accumulation of densely staining cytoplasmic inclusions. At 20 days post-injection, the morphology of the retinal pigment epithelium and photoreceptor layer of the retina in the tunicamycin-injected eyes was highly degenerate (Figs 2*d*, 3). Although many of the remaining rods in the tunicamycin-injected eyes retained some outer segment material, the rod outer segments were disoriented and were reduced in length by 73% ($P < 0.001$, $n = 3$) relative to controls. The lack of accumulated rod outer segment debris in the subretinal space and the engorged appearance of the retinal pigment epithelium indicated that phagocytosis continued after the injection of tunicamycin.

Rod outer segments exhibited a progressive and significant decrease in length as a function of post-injection time, at a rate of about 4% per day (correlation coefficient, 0.97). Even during the most advanced stages of photoreceptor degeneration, there was no significant thinning of other retinal layers (Fig. 3*A*; $P > 0.3$, $n = 3$). However, there was a progressive and selective loss of photoreceptor cells with time, resulting in nearly a 30% decrease ($P < 0.02$, $n = 3$) in the density of nuclei in the outer nuclear layer by day 20, relative to controls (Fig. 3*B*). Also, we did not observe any preferential loss of rod versus cone photoreceptors, as indicated by a relatively constant rod/cone ratio at all post-injection times.

Light microscopic autoradiography was performed after systemic labelling of the animals with ^3H -leucine to determine the incorporation of radiolabelled proteins into rod outer segment membranes, as indicated by the formation of a band of silver grains over the base of the rod outer segments. The apical

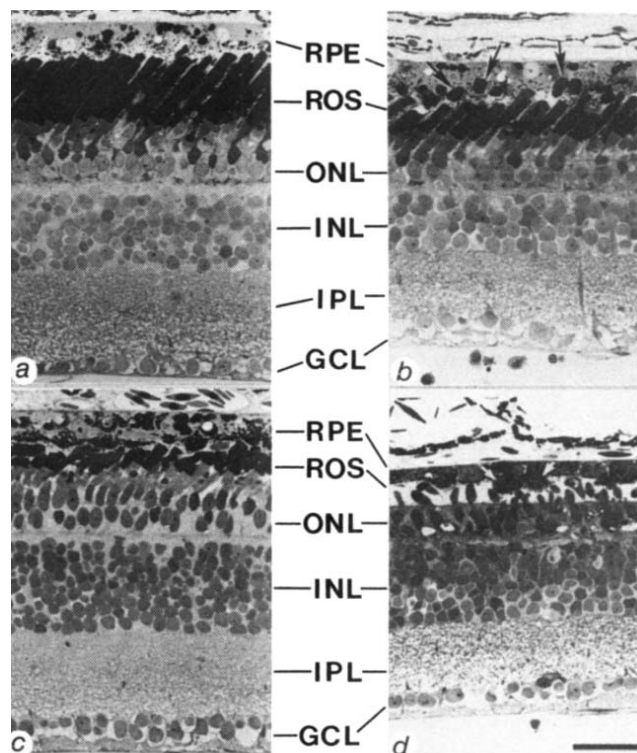


Fig. 2 Light micrographs of retinas from tunicamycin-injected eyes. Frogs were killed at (*a*) 1, (*b*) 4, (*c*) 8 and (*d*) 20 days following intraocular injection of tunicamycin, ~1 h after initial exposure to light. Fixation and processing of tissue for light microscopy was carried out as described previously¹¹. RPE, retinal pigment epithelium; ROS, rod outer segment; ONL, outer nuclear layer; INL, inner nuclear layer; IPL, inner plexiform layer; GCL, ganglion cell layer. Retinas from control eyes (not shown) exhibited morphology comparable to that shown in *a* at all post-injection times. Apparent differences in retinal cell layer thickness reflect normal regional variations and do not represent significant changes as a function of time following injection of tunicamycin (see Fig. 3*A*). Monocytes are present in the vitreous, proximal to the ganglion cell layer (*b*). Phagosomes (black arrows, *b*) of unusually large size and number are evident within the apical cytoplasm of retinal pigment epithelium cells. Migratory pigmented cells (white arrows, *d*) can be seen within the subretinal space; note the absence of rod outer segment debris in the subretinal space.

displacement of these bands was measured from the base of the rod outer segments as a function of time (Fig. 4). One day following intraocular injection, bands of silver grains were observed in retinas from both tunicamycin-injected and control eyes approximately 2.0 μ m from the base of the rod outer segments. In control eyes the displacement of these bands towards the distal tips of the rod outer segments progressed linearly at a rate of 0.92 μ m per day, in good agreement with previous determinations of rod outer segment renewal rates for this animal¹². In contrast, no further rod outer segment band displacement was observed in autoradiograms of tunicamycin-injected eyes throughout the experimental time course. These results indicate that renewal of rod outer segment membranes ceased some time within the first 4 days after the tunicamycin injection.

The photoreceptor cells are the most distal to the vitreous with respect to the other retinal cell types, and thus it is possible that the tissue concentration of tunicamycin in the photoreceptor layer was lower than in the other retinal cell layers. Therefore, it is remarkable that the degenerative changes were restricted to the photoreceptor cells; while the effects of tunicamycin on other retinal neurones were apparently negligible. The selective effect of tunicamycin on the photoreceptor cells may be due to the fact that their oxidative metabolic rate is similar to that of neoplastic cells¹³, which have been shown to be extremely sensi-

Fig. 3 Morphometric analysis of retinal cell layers as a function of time following intraocular injections with DMSO (control) or tunicamycin (TM). Measurements were made using a light microscope ($\times 25$ lens) coupled to a Zeiss Videoplan image analysis system via a video camera. Data represent the mean $\pm$ s.d. of 3 retinas for each condition; 5 random regions were analysed per retinal section and 15–20 individual measurements were performed per region. **A**, Length of rod outer segments (*a*) and thickness of the outer nuclear layer (*b*), inner nuclear layer (*c*), and the entire retinal expanse measured from the outer limiting membrane to the inner limiting membrane (*d*). Open bars, controls; shaded bars, TM. **B**, Quantitation of photoreceptor nuclei in the outer nuclear layer (ONL) (per 100 μ m linear expanse) in tunicamycin-injected eyes (●) and controls (○). Linear regression analysis indicated the rate of photoreceptor cell loss to be about 1.6% per day (correlation coefficient, 0.90).

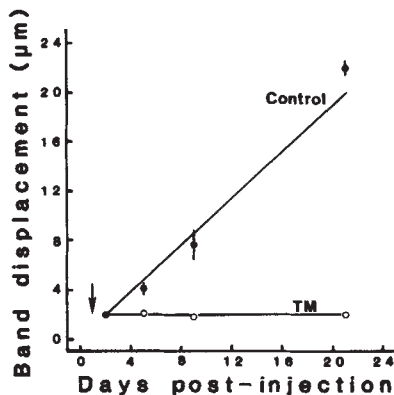
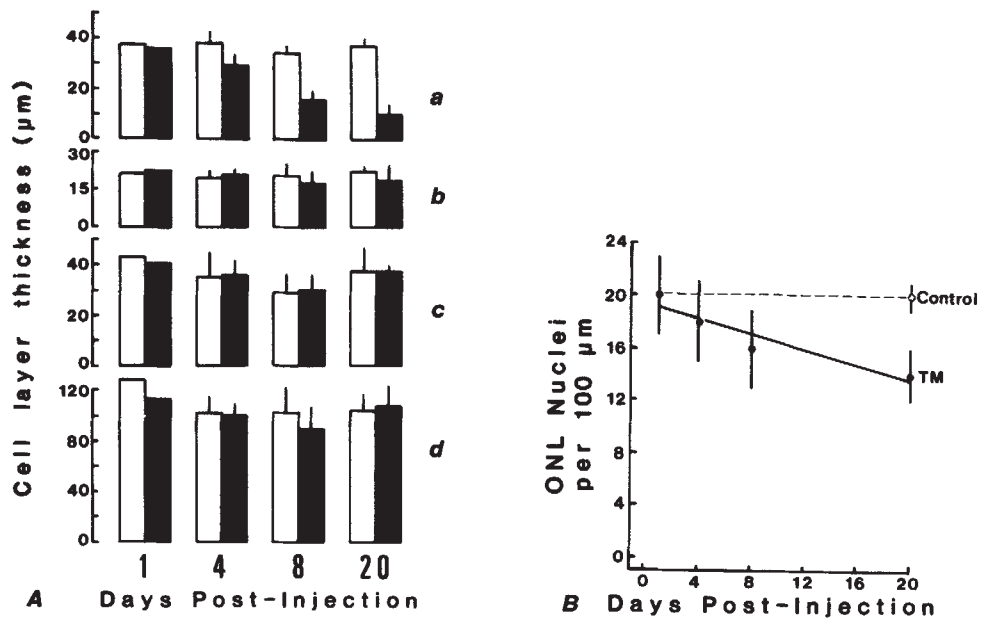


Fig. 4 Displacement of bands of silver grains in autoradiograms of rod outer segments as a function of post-injection time. **Methods:** Frogs were injected systemically with L-[4, 5- 3 H] leucine (Amersham; 130 Ci mmol $^{-1}$; 10 μ Ci per g body weight) in the dorsal lymph sack. After 24 h (indicated by arrow), intraocular injection with DMSO (●) or tunicamycin (○) was performed (see Fig. 1 legend), permitting sufficient time for the assembly of several radiolabelled disk membranes in the basal portion of the rod outer segments before the introduction of the drug. At various times (2, 5, 9 and 21 days after 3 H-leucine injection), animals were killed approximately 1 h after initial exposure to light and the eyes were processed for light microscopic autoradiography¹¹. Band displacement from the base of the rod outer segments was measured using a light microscope ($\times 100$ oil immersion lens) coupled to a Zeiss Videoplan image analysis system via a video camera. Quantitative analysis was performed as described in Fig. 2 legend. Linear regression analysis of the control (DMSO) points gave a slope of 0.92 μ m per day (correlation coefficient, 1.00), corresponding to a daily renewal equivalent to 2.2% of the length of the rod outer segment.

retinitis pigmentosa^{16–20}. A feature of these progressive blinding disorders is the selective deterioration of the photoreceptor cells, with the initial degeneration of the outer segments followed by pyknosis and cell death. The molecular etiology underlying these inherited retinal dystrophies remains unknown. These results suggest the novel possibility that a defect in the lipid-intermediate glycosylation pathway may be involved in one or more of these disorders. In this regard, we have recently obtained evidence which indicates grossly abnormal protein glycosylation in the rod photoreceptors of a human retina from a patient with an inherited chorioretinal degeneration resembling sectoral retinitis pigmentosa²¹. We are currently pursuing biochemical investigations to determine which retinal glycoproteins are affected during the tunicamycin-induced photoreceptor degeneration, as well as more detailed microscopic studies to assess potentially subtle ultrastructural changes which may occur in photoreceptors and other retinal cells during early times following intraocular injection of tunicamycin.

This work was supported by grants from the National Institutes of Health (NIH/NEI), the National Retinitis Pigmentosa Foundation (Baltimore), and the Retina Research Foundation (Houston). We thank Ms Janis Rosenthal and Mrs Mary E. Rayborn for technical assistance.

Received 24 April; accepted 25 July 1984.

1. Takatsuki, A., Kohno, K. & Tamura, G. *Agric. Biol. Chem.* **39**, 2089–2091 (1975).
2. Tkacz, J. S. & Lampen, J. O. *Biochem. biophys. Res. Commun.* **65**, 248–257 (1975).
3. Lehle, L. & Tanner, W. *FEBS Lett.* **71**, 167–170 (1976).
4. Struck, D. K. & Lennarz, W. J. in *The Biosynthesis of Glycoproteins and Proteoglycans* (ed. Lennarz, W. J.) 35–83 (Plenum, New York, 1980).
5. Tamura, G. (ed.) *Tunicamycin* (Japan Scientific Societies, Tokyo, 1982).
6. Olden, K., Parent, J. B. & White, S. *Biochim. biophys. Acta* **650**, 209–232 (1982).
7. Hargrave, P. A. in *Progress in Retinal Research* Vol. 1 (eds Osborne, N. & Chader, G.) 1–51 (Pergamon, New York, 1982).
8. Plantner, J. J., Poncz, L. & Kean, E. L. *Archs Biochem. Biophys.* **201**, 527–532 (1980).
9. Fliesler, S. J., Rayborn, M. E. & Hollyfield, J. G. *J. Cell Biol.* **97**, 413a (1983).
10. Rapp, L. M. & Basinger, S. F. *Vision Res.* **22**, 1097–1103 (1982).
11. Hollyfield, J. G., Rayborn, M. E., Sarthy, P. V. & Lam, D. M. K. *J. comp. Neurol.* **188**, 587–598 (1979).
12. Young, R. W. *Invest. Ophthalmol. vis. Sci.* **15**, 700–725 (1976).
13. Graymore, C. N. in *Biochemistry of the Eye* (ed. Graymore, C. N.) 645–735 (Academic, New York, 1970).
14. Yamada, K. M. & Olden, K. in *Tunicamycin* (ed. Tamura, G.) 119–144 (Japan Scientific Societies, Tokyo, 1982).
15. Fliesler, S. J., Tabor, G. A. & Hollyfield, J. G. *Expl Eye Res.* (in the press).
16. Kolb, H. & Gouras, P. *Invest. Ophthalmol. vis. Sci.* **13**, 487–498 (1974).
17. Szamier, R. B. & Berson, E. L. *Invest. Ophthalmol. vis. Sci.* **16**, 947–962 (1977).
18. Szamier, R. B., Berson, E. L., Klein, R. S. & Meyers, S. *Invest. Ophthalmol. vis. Sci.* **18**, 145–160 (1979).
19. Rayborn, M. E., Moorhead, L. C. & Hollyfield, J. G. *Ophthalmology* **89**, 1441–1454 (1982).
20. Bunt-Milam, A. H., Kalina, R. E. & Pagon, R. A. *Invest. Ophthalmol. vis. Sci.* **24**, 458–469 (1983).
21. Hollyfield, J. G., Frederick, J. M., Tabor, G. A. & Ulshofer, R. J. *Ophthalmology* **91**, 191–196 (1984).

tive to very low concentrations of tunicamycin¹⁴. In addition, the relative activity of the lipid-intermediate pathway for glycoprotein biosynthesis in the retina is greatest in the photoreceptor layer¹⁵. Therefore, tunicamycin might be expected to exert its inhibitory effects most prominently in the photoreceptor cells. Alternatively, the relative affinity of the uptake system for tunicamycin may be higher in photoreceptors than in other retinal cells.

These results are reminiscent of the kinds of pathological features observed in human rod-cone dystrophies, including

Recombinant vaccinia virus primes and stimulates influenza haemagglutinin-specific cytotoxic T cells

J. R. Bennink*, J. W. Yewdell*, G. L. Smith†, C. Moller* & B. Moss†

* The Wistar Institute, 36th Street at Spruce, Philadelphia, Pennsylvania 19104, USA

† Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, Maryland 20205, USA

The ability of vaccinia virus to accept and express cloned genes¹⁻³ encoding immunologically important proteins of unrelated viruses⁴⁻⁹ and malarial parasites¹⁰ has suggested a novel approach to the development of live vaccines. Vaccinia virus recombinants retain infectivity and stimulate synthesis of specific antibodies to the cloned gene products in vaccinated animals. Moreover, animals inoculated with recombinants expressing the influenza virus haemagglutinin (HA)⁷, the hepatitis B virus surface antigen¹¹, and type 1 herpesvirus glycoprotein D⁸ were protected against subsequent challenge with the corresponding virus. For maximal effectiveness, vaccines should produce cellular as well as humoral immunity. We now report that a vaccinia virus recombinant, expressing the influenza HA, primes and stimulates a specific murine cytotoxic T-lymphocyte (CTL) response. Histocompatible cells infected with this recombinant also serve as targets for CTLs. These properties make vaccinia virus a unique tool for studying cell-mediated immunity and enhance the attractiveness of this vector for production of live vaccines.

The construction of the vaccinia virus recombinant H2-VAC (formerly referred to as vInf1) expressing the HA gene of influenza virus A/Jap/305/57 (H2N2) (JAP) has been described previously⁷. The HA gene was inserted next to a vaccinia virus transcriptional start site so that the authentic translational initiation and termination codons of the influenza gene were used. Tissue culture cells infected with the recombinant virus synthesized an HA polypeptide of the correct size that was glycosylated and transported to the cell surface^{7,9}. The latter feature suggested that histocompatible cells infected with the recombinant virus might serve as targets for influenza virus-specific CTLs. To test this directly, P815 cells (a mastocytoma line derived from DBA/2 (*H-2^d*) mice) were infected and used as targets in a ⁵¹Cr-release assay. CTLs derived from BALB/c (*H-2^d*) mice previously inoculated with JAP and restimulated with JAP-infected autologous splenocytes were able to lyse both JAP- and H2-VAC-infected cells (Fig. 1). In contrast, only background levels of cell lysis were observed using targets infected with wild-type vaccinia virus (Fig. 1). Control experiments verified that the latter were effectively lysed by CTLs prepared from vaccinia virus-infected mice (not shown). These results confirm our previous evidence that the HA is expressed on the surface of cells infected with recombinant vaccinia virus and further demonstrate that the HA can interact with class I major histocompatibility complex molecules in a manner sufficient for recognition by CTLs.

We next examined the ability of the recombinant vaccinia virus to restimulate *in vitro* splenocytes derived from BALB/c mice primed 3 weeks earlier by infection with JAP. Figure 2 shows that CTLs stimulated by H2-VAC were able to efficiently lyse cells infected with JAP or H2-VAC itself but not cells infected with an influenza B virus or wild-type vaccinia virus. As anticipated, we repeatedly found no detectable anti-influenza virus CTL activity when wild-type vaccinia virus was used for stimulation (not shown).

Further experiments demonstrated that the recombinant vaccinia virus could also prime for an influenza virus-specific CTL response. BALB/c mice were inoculated with recombinant or wild-type vaccinia virus; after 4 weeks, splenocytes were stimu-

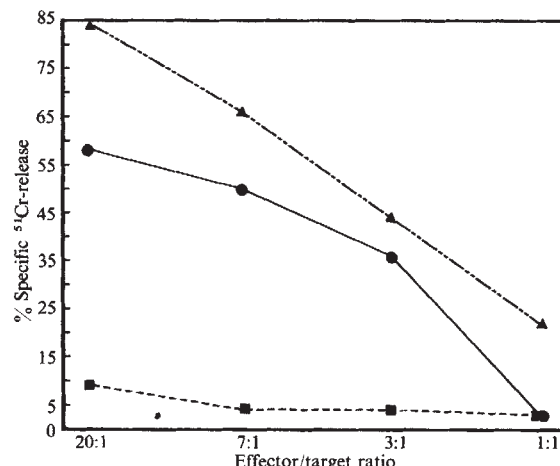


Fig. 1 Lysis of H2-VAC-infected target cells (—●—) by anti-influenza virus-specific CTLs. BALB/c splenocytes for *in vitro* stimulation were obtained from mice 2–4 weeks after intraperitoneal immunization with allantoic fluid containing 200 haemagglutinating units of JAP. Primed spleen cells (5×10^7) were cultured for 6 days in the presence of 2.5×10^7 autologous spleen cells infected with JAP at a multiplicity of 100 egg infectious doses (EID) per cell, then tested in a standard ⁵¹Cr-release assay. P815 cells were infected with vaccinia virus WR isolate (—■—) or recombinant H2-VAC at a multiplicity of 10–100 plaque-forming units (PFU) per cell or with JAP (—●—) at 10–100 EID per cell for 1 h¹⁸. After washing, cells were incubated for 6 h (JAP infection) or 15 h (vaccinia infection), then incubated for 1 h with ⁵¹Cr at 37 °C, washed twice and dispensed into wells of microtitre plates (10^4 cells per well). Effector cells from *in vitro* cultures were added at various ratios, and after 4 h incubation at 37 °C, one-half of the culture supernatant was removed for determination of the concentration of released ⁵¹Cr by γ counting. Per cent specific release was determined as follows: experimental release – spontaneous release (no CTLs) divided by total release (in the presence of detergent) – spontaneous release. Spontaneous release values were: JAP, 28%; H2-VAC, 14%; and wild-type vaccinia virus, 18%.

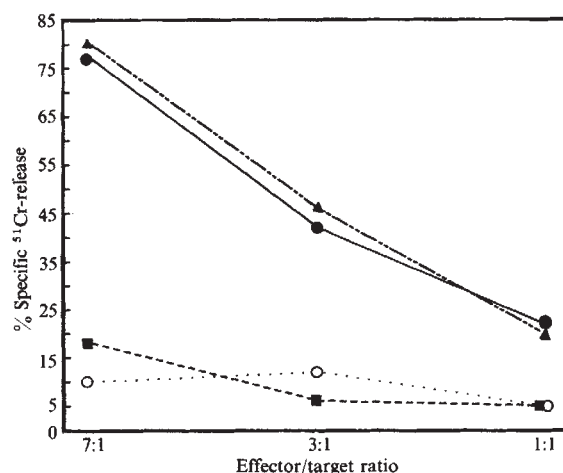


Fig. 2 Stimulation of JAP-primed CTLs by H2-VAC. Splenocytes from BALB/c mice primed with JAP were cultured for 6 days in the presence of BALB/c spleen cells infected with the recombinant H2-VAC virus (10 PFU per cell). ⁵¹Cr-release assays were performed on P815 cells as described in Fig. 1 legend. Spontaneous release values were: JAP (—●—), 18%; H2-VAC (—■—), 16%; wild-type vaccinia virus (—○—), 14%; and B/Lee (· · · ○ · · ·), 39%.

lated *in vitro* with JAP. Figure 3a shows that the recombinant virus was able to prime for an influenza virus-specific CTL response, whereas influenza virus-specific priming did not occur when mice were inoculated with wild-type vaccinia virus (Fig. 3b). Note that CTLs induced by H2-VAC priming and JAP secondary stimulation lysed cells infected with H2-VAC and

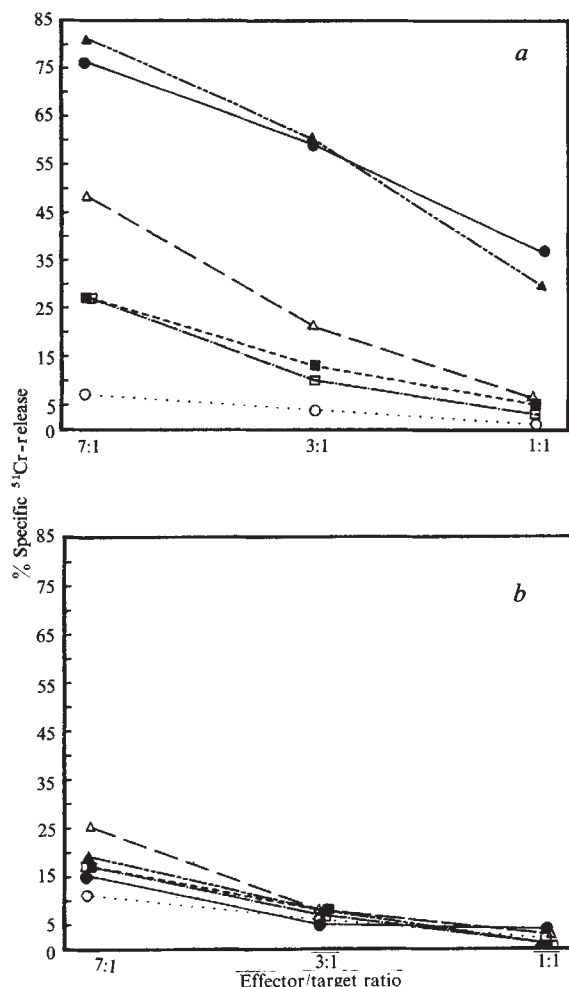


Fig. 3 *a, b*, Priming for an anti-JAP-specific CTL response by H2-VAC. Splenocytes were obtained from BALB/c mice 2–4 weeks after intravenous immunization with 10^7 PFU of the vaccinia recombinant H2-VAC (*a*) or the wild-type WR isolate of vaccinia virus (*b*). Primed splenocytes were restimulated *in vitro* with JAP-infected splenocytes. Target cells were prepared as described in Fig. 1 legend except that, after infection, all targets were incubated for 6 h before labelling with ^{51}Cr . Spontaneous release value for all targets was <13%. Targets: ●—, H2-VAC; ■—, wild type vaccinia virus; —△—, JAP; ○··, B/Lee; —△—, PR8; —□—, HK.

JAP to a similar extent—this demonstrates that at least a significant proportion of these cells are specific for the HA. It is also apparent that cells infected with influenza A viruses of other subtypes [A/PR/8/34 (H1N1) (PR8) and A/HK/107/68 (H3N2) (HK)] were lysed at levels significantly greater than observed with B/Lee-infected cells. This could be due to cross-reactive recognition of heterosubtypic haemagglutinins by anti-JAP HA CTLs¹², in which case the better killing of PR8-infected cells could reflect the higher amino acid homology between JAP and PR8 haemagglutinins compared with JAP and HK haemagglutinins¹³. Alternatively, the observed cross-reactive lysis could be due to CTLs specific for other viral components whose *in vitro* stimulation depends on the presence of HA-specific T-helper cells primed by H2-VAC inoculation. In either case, it is important that priming for CTLs able to lyse cells infected with influenza viruses of H1N1 and H3N2 subtypes was not observed consistently between experiments; this, and the relatively small cross-reactive response, suggests that it may be due to stimulation of relatively infrequent T-cell clonotypes. Finally, as anticipated from the preceding experiments, virus priming and stimulation with H2-VAC alone resulted in the generation of JAP-specific CTLs (Fig. 4).

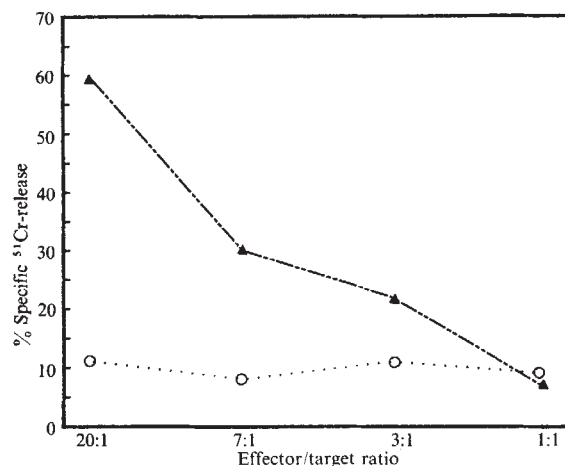


Fig. 4 Generation of anti-JAP-specific CTLs by priming and stimulation with H2-VAC. Splenocytes from BALB/c mice primed with H2-VAC were cultured for 6 days in the presence of BALB/c spleen cells infected with H2-VAC and tested for cytotoxic activity. Target cells were prepared as described in Fig. 1 legend. Spontaneous release values were: JAP (—△—), 18%; B/Lee (·····), 35%.

These data are of practical importance in two respects. First, we have shown that cells infected with a recombinant vaccinia virus can serve as excellent targets for CTLs specific for the product of the inserted gene. This, coupled with a wide host cell range, makes recombinant vaccinia virus an excellent tool for determining which viral antigens are recognized by CTLs, a question which remains unanswered for almost all viruses. The ability of vaccinia virus to accommodate large amounts of inserted DNA¹⁴ means that it would even be possible to include multiple viral genes or viral and major histocompatibility complex genes in the same vaccinia virus vector. Second, regarding the potential use of recombinant vaccinia viruses as live vaccines, we have shown that these vaccines will probably efficiently prime recipients for secondary CTL responses. Although the role of CTLs in host defence against viral infection remains to be precisely defined¹⁵, present evidence suggests that they reduce replication and dissemination of at least some viruses^{16,17}. This property of recombinant vaccinia viruses may offer a significant advantage over subunit or inactivated vaccines, which do not, in general, prime for anti-viral CTL responses.

This work was supported by NIH grants AI 14162, AI 20338, AI 13989 and NS 11036.

Received 11 May; accepted 15 August 1984.

1. Panicali, D. & Paoletti, E. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4927–4931 (1982).
2. Mackett, M., Smith, G. L. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7415–7419 (1982).
3. Mackett, M., Smith, G. L. & Moss, B. *J. Virol.* **49**, 857–864 (1984).
4. Smith, G. L., Mackett, M. & Moss, B. *Nature* **302**, 490–495 (1983).
5. Smith, G. L., Mackett, M. & Moss, B. in *UCLA Symp. molec. cell. Biol. new Ser.* (eds Hamer, D. & Rosenberg, M.) 543–554 (Liss, New York, 1983).
6. Panicali, D., Davis, S. W., Weinberg, R. L. & Paoletti, E. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5364–5368 (1983).
7. Smith, G. L., Murphy, B. R. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7155–7159 (1983).
8. Paoletti, E., Lipinskas, B. R., Samsonoff, C., Mercer, S. & Panicali, D. *Proc. natn. Acad. Sci. U.S.A.* **81**, 193–197 (1984).
9. Smith, G. L., Mackett, M., Murphy, B. R. & Moss, B. in *Modern Approaches to Vaccines* (eds Chanock, R. & Lerner, L.) (Cold Spring Harbor Laboratory, New York, in the press).
10. Smith, G. L., Godson, B. N., Nussenzweig, V., Nussenzweig, R. S. & Moss, B. *Science* **224**, 397–399 (1984).
11. Moss, B., Smith, G. L., Gerin, J. L. & Purcell, R. H. *Nature* **311**, 67–69 (1984).
12. Braciale, T. J., Braciale, U. L., Henkel, T. J., Sambrook, J. S. & Gething, M. J. *Exp. Med.* **159**, 341–354 (1984).
13. Winter, G., Fields, S. & Brownlee, G. G. *Nature* **292**, 72–75 (1981).
14. Smith, G. L. & Moss, B. *Gene* **25**, 21–28 (1983).
15. Zinkernagel, R. M. & Doherty, P. C. *Adv. Immun.* **27**, 51–77 (1979).
16. Yap, K. L., Ada, G. L. & McKenzie, I. F. C. *Nature* **273**, 238–239 (1978).
17. Lin, Y. & Askanas, B. A. *J. exp. Med.* **154**, 225–234 (1981).
18. Yewdell, J. W., Frank, E. & Gerhard, W. *J. Immun.* **126**, 1814–1819 (1981).

Transposition without duplication of infecting bacteriophage Mu DNA

Rasika M. Harshey

Research Institute of Scripps Clinic,
Department of Molecular Biology, 10666 North Torrey Pines Road,
La Jolla, California 92037, USA

Most models of DNA transposition invoke replication of the transposable element¹, but it is not clear whether a 'co-integrate' is an obligatory intermediate in the pathway leading to the production of simple insertions during transposition. Such an intermediate can be accounted for only by a replicative transposition scheme. Bacteriophage Mu is a temperate phage that can either lysogenize or lyse its host², and it encodes at least two modes of transposition as judged by the end-products generated by the process. During the lytic development of the integrated prophage, co-integrates are the predominant end-products³; transposition is coupled to replication during this phase⁴. A small number of simple insertions are also produced during the lytic growth³, but during transposition from the infecting phage into the host chromosome, simple insertions are the main end-products⁵. Conditions can be found where the choice between the two kinds of end-products depends on a delicate balance between the essential transposition functions encoded by Mu⁶. Experiments^{7,8} have suggested that the simple insertions which arise during transposition from the infecting phage may do so without Mu DNA replication. Here I demonstrate using an infecting phage with completely methylated DNA, a *dam*⁻ (DNA adenine methylase) host and a combination of restriction enzymes that can cut either fully methylated or unmethylated DNA but not hemi-methylated DNA, that transposition of the phage DNA into the host chromosome does not involve a duplication of its DNA. This result may also have significance for other transposons⁹ that do not appear to go through a co-integrate intermediate during transposition¹⁰.

The *dam* gene of *Escherichia coli* codes for an enzyme that methylates the N-6 position of deoxyadenosine residues in DNA¹¹. The methylated sequence, GATC, is cleaved by the

restriction enzyme *DpnI*. Another enzyme, *MboI*, cleaves the same sequence, but only if the sequence is unmethylated; however neither enzyme will cleave a hemi-methylated sequence¹². I have used this property to design a strategy for investigating the state of replication of infecting Mu DNA, when it transposes into its host chromosome.

Phage [(Mu *cts* 62 445-5, *gin*⁻(G+)]¹³ were methylated by growing them in a strain of *E. coli* containing the cloned *dam* gene (p *dam* 118)¹⁴. The DNA was digested with *DpnI* (data not shown) to show that it was completely methylated. These phage were then used to infect a *dam*⁻ *E. coli* host. At various times after infection, total *E. coli* DNA was isolated and checked for integration of Mu (Fig. 1A). Integration can be detected 8 min after infection (the exact time of detection of integration varies, being as early as 4 min in some experiments). Chromosomal DNA was then separated from unintegrated Mu DNA by non-equilibrium centrifugation on a sodium chloride gradient (Fig. 1A, lane *d*). Digestion of the integrated Mu DNA with *HindIII* showed that the two central *HindIII* fragments present in the phage DNA still remained, but the terminal *HindIII* fragments present in the phage DNA had disappeared, to be replaced by a smear, as would be expected from molecules that had transposed from the phage DNA into many different sites in the chromosome (Fig. 1B). The integrated DNA was also digested with *EcoRI* which releases larger terminal fragments. Again, only the central fragment could be detected (data not shown). Using these criteria, contamination from free phage DNA was judged to be <0.5%.

DNA isolated 8 min post-infection was digested with *MboI* and *DpnI* respectively (Fig. 2a, b). Integrated Mu DNA is resistant to the action of *MboI* but completely digested by *DpnI*, showing that no replication of the DNA had occurred during its transposition. However, this DNA is capable of further rounds of transposition as seen by the digestion pattern 15 min post-infection. Figure 2c shows that some of the integrated DNA is resistant to both *MboI* and *DpnI* whereas some has become susceptible to *MboI*, indicating that a fraction of the integrated Mu DNA has gone through one round of replication (become hemi-methylated and hence resistant to both enzymes) and even two rounds (become unmethylated and hence sensitive to *MboI*). By 30 min post-infection, most of the integrated Mu

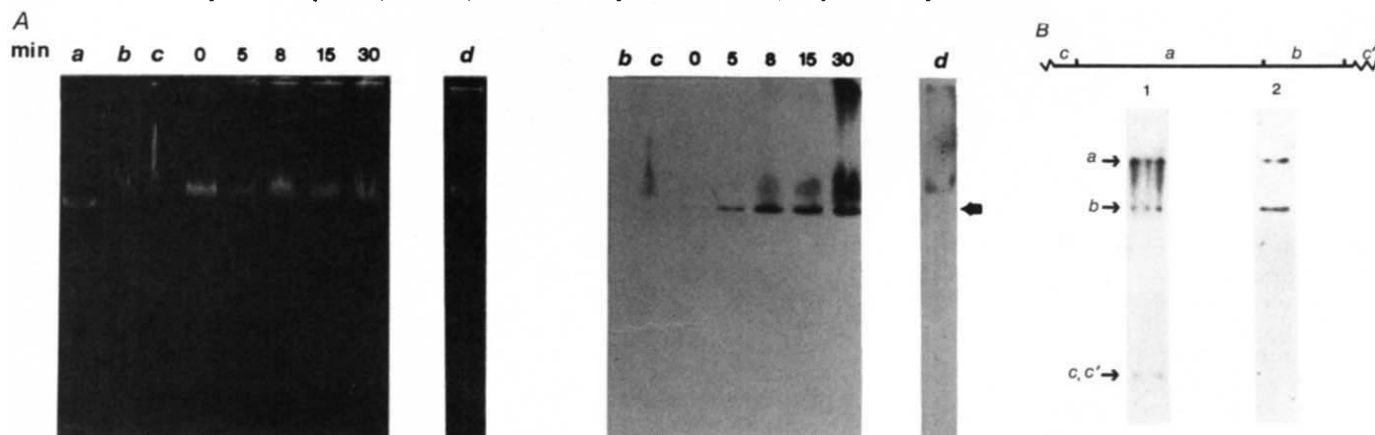


Fig. 1 Chromosomal integration of infecting Mu DNA. **A**, Total *E. coli* DNA (GM119) isolated at various times after phage infection was run on a 0.35% agarose gel, transferred to nitrocellulose and probed with radiolabelled Mu DNA. The left panel shows the ethidium-bromide stained fluorescence and the right panel the autoradiogram of the gel. Lane *a*, DNA isolated from phage; lane *b*, control *E. coli* DNA; lane *c*, DNA from a Mu lysogen. 0 to 30 indicate the time in minutes after infection at which total *E. coli* DNA was isolated. Lane *d*, DNA from the 8 min time point, separated from free phage DNA as described below. The arrow indicates the position of free phage DNA. **B**, Schematic diagram showing the distribution of *HindIII* sites in Mu 445-5. ~ represents heterogeneous host sequences. *a*, *b* Represent the internal fragments and *c* and *c'* the terminal fragments. *c* and *c'* are similar in size and band as a doublet. Lane 1, the *HindIII* digestion pattern of DNA from 445-5 phage, and lane 2 that from integrated Mu DNA 8 min post-infection.

Methods: Cells (GM119, *dam*, *dcm*-6) were grown to a cell density of 2×10^8 per ml at 37 °C. They were pelleted, resuspended on ice in buffer containing 20 mM Tris, 200 mM NaCl, 2 mM CaCl₂, 20 mM MgCl₂, pH 7.5. Phage were added at a multiplicity of infection (MOI) of 5, mixed gently for 1–2 min and diluted into pre-warmed broth at 37 °C. Cells were left shaking at 37 °C and at various times aliquots were withdrawn and total DNA isolated as described elsewhere⁴. Integrated Mu DNA was separated from the free phage DNA by running not more than 100 µg of total DNA over 10–30% NaCl gradients in 10 mM Tris, 1 mM EDTA pH 7.5, spun in a SW41 rotor, 25 °C, 37,000 r.p.m. for 3 h. High molecular weight DNA pelleted; it was recovered, redigested with pronase, re-extracted with phenol and precipitated with ethanol.

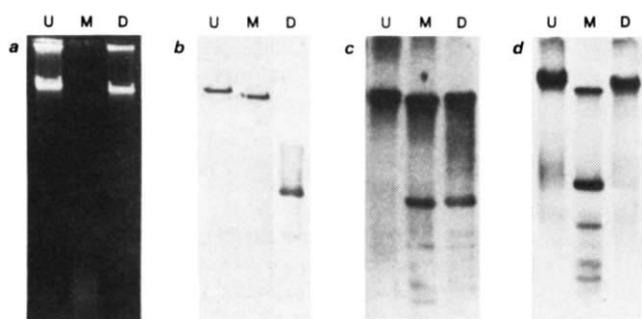


Fig. 2 *MboI* and *DpnI* digestion pattern of integrated Mu DNA. Agarose gels (0.5%) were run as in Fig. 1. *a*, Ethidium-bromide stained fluorescence of DNA 8 min post-infection. The DNA was digested with *MboI* (lane M) (New England Biolabs) or *DpnI* (lane D) (Boehringer Mannheim) according to manufacturer's specifications. Control lanes contain undigested DNA (lane U). *b*, Autoradiogram of the same gel. *c*, Autoradiogram of DNA isolated 15 min post-infection and treated as in *a*. *d*, Autoradiogram of DNA isolated 30 min post-infection and treated as in *a*.

DNA is resistant to *DpnI* (Fig. 2*d*). These results are consistent with observations that by 30 min after phage induction, Mu DNA is actively undergoing replicative transposition^{4,15}.

Although most of the integrated Mu DNA appears to be digested with *DpnI* 8 min after infection, showing that integration had occurred without replication, it would nevertheless be expected that 170 base pairs (bp) or ~0.46% of Mu DNA would still remain linked to *E. coli* DNA, as the first *DpnI* sites occur 56 bp from the left end and 114 bp from the right end of Mu. A densitometric tracing of the autoradiogram in Fig. 2*b* indicates, however, that about 4% of the integrated DNA remains resistant to *DpnI* (data not shown). There are several possible explanations of this result, the simplest being that some of the integrated molecules had already entered the subsequent pathway of replicative transposition. To eliminate this possibility, infections were performed into a *dam*⁻ host strain carrying *himA*Δ82, a mutation which allows infecting Mu to integrate normally but which does not support the subsequent lytic growth of the phage¹⁶. In this experiment, integration could be detected as early as 4 min post-infection. The integrated DNA, as in the previous experiment, was susceptible to *DpnI*, but remained so even 15 min post-infection (Fig. 3). However, 4–8% of the DNA still remained resistant to the enzyme at every time point studied. To test the possibility that this fraction represented a subpopulation of molecules that transposed using the replicative mode, the DNA was double-digested with *MboI* and *DpnI* (Fig. 3, 15 min, lane M/D). If such a subpopulation existed, this fraction would have been hemi-methylated for the entire length of Mu, and would have remained in the high molecular weight region of the gel. However, the *DpnI* resistant material in lane D no longer remained in the high molecular weight DNA fraction in lane M/D, indicating that it represents partial replication of some or all the integrated Mu molecules. Such replication could be unrelated to the transposition process itself; for example, it could occur by passive replication of the integrated Mu DNA by the *E. coli* replication machinery, or it could be an essential feature of the transposition mechanism which generates simple insertions. It is known, for example, that Mu integration generates a 5-bp duplication at the target site^{17,18}, which is proposed to arise by a staggered break at the site of insertion and which is subsequently repaired¹⁹. Some degree of nonspecificity in this process might lead to repair synthesis extending beyond the 5 bp and into Mu sequences. I favour this latter explanation because if the former had been the case, one would have observed an increase in the *DpnI* resistant material with time, and this did not occur.

Our results indicate that Mu DNA transposition from an infecting phage into the chromosome of its host occurs without a duplication of its genome, in sharp contrast to its subsequent

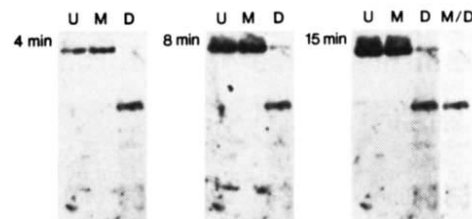


Fig. 3 *MboI* and *DpnI* digestion pattern of integrated Mu DNA isolated from a *dam*⁻ strain carrying *himA*Δ82 (RH300). Agarose gels (1%) were run as in Fig. 1. Lanes U, M and D refer to undigested controls, *MboI* and *DpnI* cleaved samples, respectively. The patterns for 4 min, 8 min and 15 min post-infection are shown.

transposition events which are replicative⁴. This then raises questions of what factors govern the choice of the transposition pathway, either simple insertions or co-integrates. One obvious candidate is the physical state of the DNA (although phage DNA is linear, after infection it can be isolated as a non-covalently-closed circular DNA-protein complex²⁰). Another possibility is a differential regulation of transposition functions⁶. It is possible that the state of DNA in turn influences the expression of specific proteins.

During the transposition of Tn3, the occurrence of an intermediate co-integrate state which is resolved by a site-specific recombination system to produce simple insertions²¹, led to the proposal that co-integrates are intermediates in all DNA transpositions²². The results reported here prove that is not so. They do not, however, rule out the possibility that there are other circumstances in which simple insertions could arise by a replicative pathway¹.

When determining the molecular mechanism of a non-replicative transposition reaction, the fate of the donor molecule on which the transposable element resides (in the case of infecting Mu, the heterogeneous *E. coli* sequences at its ends) must be considered. The donor molecule may be repaired after the element has 'jumped' out, or may not be repaired. If unrepaired, it would result in the destruction of the donor (either event would be irrelevant for the survival of the infecting Mu phage)⁹. Such a suicidal, non-replicative mode of DNA transposition may explain the occurrence of chromosomal breaks associated with the movement of some transposable elements²³. One may then raise general questions about the evolutionary significance of replicative versus non-replicative mechanisms for DNA transposition and why different movable elements make the choices they do.

I dedicate this paper to the memory of Ahmad Bukhari. I thank M. Jayaram for many discussions and Richard Lerner for his support. This work was supported by grants from the NSF (PCM8203174) and the NIH (GM33247-01).

Received 14 May; accepted 3 August 1984.

1. Bukhari, A. I. *Trends biochem. Sci.* **6**, 56–60 (1981).
2. Bukhari, A. I. *A. Rev. Genet.* **10**, 389–411 (1976).
3. Chaconas, G., Harshey, R. M., Sarvetnick, N. & Bukhari, A. I. *J. molec. Biol.* **150**, 341–359 (1981).
4. Ljungquist, E. & Bukhari, A. I. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3143–3147 (1977).
5. Chaconas, G., Kennedy, D. L. & Evans, D. *Virology* **128**, 48–59 (1983).
6. Harshey, R. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2012–2016 (1983).
7. Leibart, J. C., Ghelardini, P. & Paolozzi, L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4362–4366 (1982).
8. Akroyd, J. E. & Symonds, N. *Nature* **303**, 84–86 (1983).
9. Berg, D. E. in *DNA Insertion Elements, Plasmids and Episomes* (eds Bukhari, A. I., Shapiro, J. A. & Adhya, S. L.) 205–212 (Cold Spring Harbor Laboratory, New York, 1977).
10. Kleckner, N. A. *Rev. Genet.* **15**, 341–404 (1981).
11. Marinus, M. G. & Morris, R. N. *J. Bact.* **114**, 1143–1150 (1973).
12. Vovis, G. F. & Lacks, S. *J. molec. Biol.* **115**, 525–538 (1977).
13. Chow, L. T., Kahmann, R. & Kamp, D. *J. molec. Biol.* **113**, 591–609 (1977).
14. Brooks, J. E., Blumenthal, R. M. & Gigeras, T. R. *Nucleic Acids Res.* **11**, 837–851 (1983).
15. Razzaki, T. & Bukhari, A. I. *J. Bact.* **122**, 437–442 (1975).
16. Miller, H. I. & Friedman, D. I. *Cell* **20**, 711–719 (1980).
17. Allet, B. *Cell* **16**, 123–129 (1979).
18. Kahmann, R. & Kamp, D. *Nature* **280**, 247–250 (1979).
19. Grindley, N. D. F. & Sherratt, D. J. *Cold Spring Harb. Symp. quant. Biol.* **43**, 1257–1261 (1978).
20. Harshey, R. M. & Bukhari, A. I. *J. molec. Biol.* **167**, 427–441 (1983).
21. Gill, R., Heffron, F., Dougan, G. & Falkow, S. *J. Bact.* **136**, 742–756 (1978).
22. Shapiro, J. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1933–1937 (1979).
23. McClintock, B. *Cold Spring Harb. Symp. quant. Biol.* **16**, 13–47 (1951).

Effect of protein packing structure on side-chain methyl rotor conformations

A. A. Kossiakoff*†‡ & S. Shteyn†

* Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143, USA

† Department of Biology, Brookhaven National Laboratory, Upton, New York 11973, USA

Protein molecules undergo a series of conformational fluctuations ranging in degree from atomic vibrations to transient denaturation, even in physiological conditions^{1,2}. The rotational motions of amino acid side chains form an important subset of the types of fluctuation a protein can undergo. NMR³⁻⁷ and molecular dynamics⁸⁻¹⁰ have shown that methyl groups in proteins are not held in fixed positions, but spin rapidly around their rotor axes. The question then arises as to whether methyl groups in proteins predominantly adopt the 'staggered' conformation, favoured by the intrinsic barrier to rotation of these groups, or whether cooperative packing effects in the folded protein perturb the average configurations to higher torsional energy. We report here an investigation of the rotational conformations of the methyl groups of aliphatic side chains in the protein crambin by neutron diffraction. We find that in the time-averaged structure of this protein, the majority of methyl rotors adopt the staggered conformation. This is consistent with rotation being a quantized event consisting of rapid reorientations of $\sim 120^\circ$ steps to positions of highest stability. The fact that the local environment does not dictate the low energy state of methyl groups suggests that within the seemingly close-packed interior structure of a protein, mutual packing accommodation occurs as a consequence of the inherent flexibility and small packing defects in protein structures.

Neutron diffraction is an ideal method for investigating methyl group conformation because it allows direct observation of hydrogen atom positions in neutron density maps, and also measurement of the time-averaged structure and thus identification of the lowest energy conformer. Previously, neutron diffraction of trypsin demonstrated that high-resolution data and a well refined phasing model are a prerequisite of any type of detailed methyl rotor analysis¹¹. Crambin, a small protein consisting of a single polypeptide chain of 46 amino acid residues^{12,13}, with 27 terminal methyl groups, therefore makes an ideal candidate for this type of study. Data are obtained from a neutron structure analysis at 1.4 Å resolution. Details of the experimental procedures have been reported previously¹⁴.

Figure 1 shows examples of methyl group neutron densities from the crambin analysis. These densities are taken from a difference map with the density contribution of the methyl carbon subtracted (see Fig. 1). It is clear that for most cases, unambiguous rotor orientations can be assigned from maps of this quality. The rotor density profiles can be displayed graphically by density plots (Fig. 2). These profiles plot the density sampled every 10° around the rotation axis at the stereochemically appropriate location for a possible methyl hydrogen position. Most profiles show a pattern for hydrogen positions every 120° around the rotation axis, which indicates a high degree of order. However, there are a significant number of cases that are disordered by crystallographic criteria. This does not necessarily indicate rotational disorder alone, but could be a function of the inherent vibrational motion of the parent methyl carbon itself and hence of the entire methyl group.

The least well ordered methyl group pictured is Ala 45 CB (Fig. 1d). Its corresponding rotor density plot (Fig. 2, symbol +), however, shows a symmetrical profile from which a 'preferred' orientation can be identified. We find in several instances that although the fine structures of this type of curve showed varying degrees of asymmetry (presumably due to residual phas-

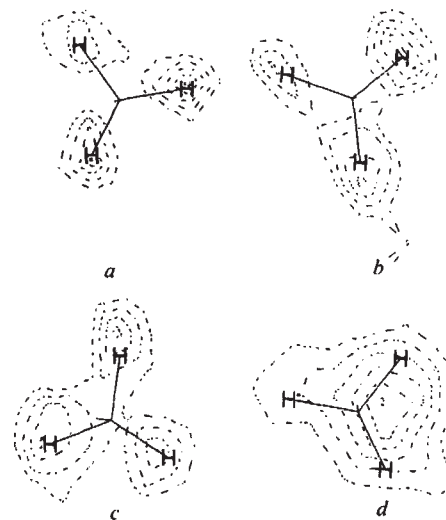


Fig. 1 Sections of the neutron difference Fourier map showing methyl hydrogen densities for several representative methyl groups. The groups shown are: a, Ala 24; b, Thr 21; c, Thr 28; and d, Ala 45. Because of the unique scattering characteristics of neutrons by hydrogen atoms, hydrogen atom positions appear in neutron density maps as negative peaks. No phase information about the methyl hydrogens was included in the model; therefore, hydrogens appear in the difference map at their true positions but at reduced density ($\approx 1/2$ weight). This crambin analysis represents the first time that methyl hydrogens have ever been observed at atomic resolution in a Fourier density map.

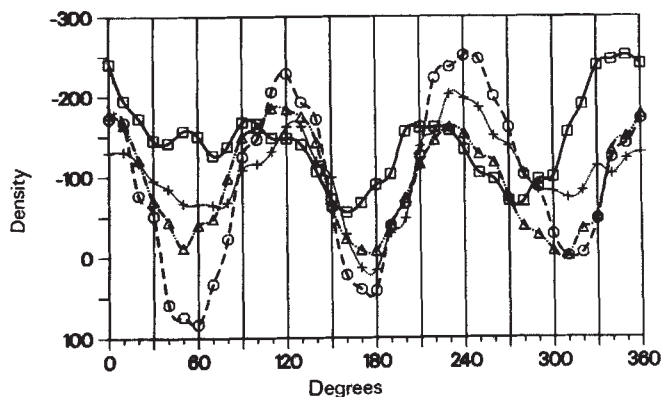


Fig. 2 Examples of representative methyl rotor plots calculated by sampling the map densities in incremental steps around the rotor axes. □, Val 15 CGI; △, Thr 21; ○, Ala 27; +, Ala 45. These plots are derived by calculating a series of possible methyl hydrogen positions at points differing by a rotational increment of 10° and then sampling the observed density at each point in the difference map. Many of the rotor plots have a 3-fold symmetrical pattern, which is due to real structural features since there is no inherent bias in the phasing to produce the 3-fold character. The torsion angle for a staggered conformation is defined here as $0^\circ, 20^\circ, 240^\circ$ (this corresponds to the usual convention of $60^\circ, 180^\circ, 300^\circ$).

ing error), the overall shapes of the profiles generally indicated a systematic trend of peaks and valleys. The rationale in assigning a specific orientation to a rotor is because the peaks and valleys usually appear at intervals of 120° , irrespective of their relative amplitudes. Since no bias was introduced into the maps, the probability of this pattern occurring by chance is remote.

The identification of the highest populated rotor orientation of each methyl group can be assigned most accurately by a 3-fold averaging procedure. In theory, the shape of the resulting

† Present address: Genentech, Inc., 460 Point San Bruno Blvd, South San Francisco, California 94080, USA.

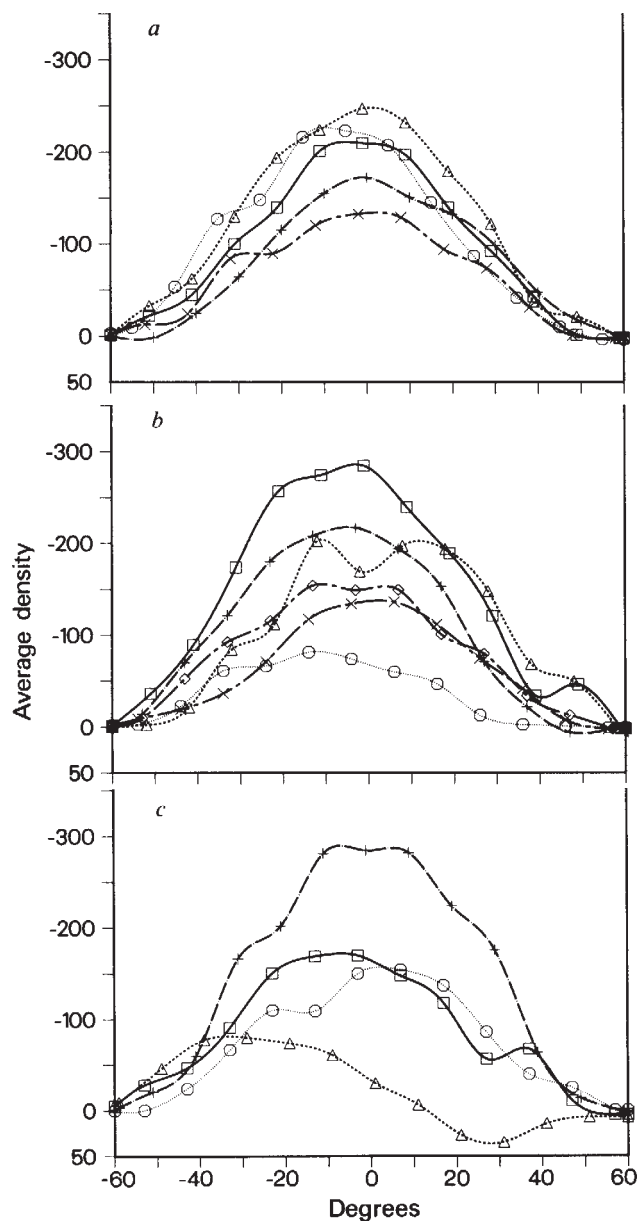


Fig. 3 Rotationally averaged rotor plots for alanine (a), threonine (b) and valine (c). a, Ala CB: $\square$ 9, $\circ$ 24, $\triangle$ 27, $+$ 38, $\times$ 45; b, Thr CG2: $\square$ 1, $\circ$ 2, $\triangle$ 21, $+$ 28, $\times$ 30, $\diamond$ 39; c, Val CG1: $\square$ 8, $\triangle$ 15; CG2: $\circ$ 8, $+$ 15. Plots were calculated by averaging densities at points separated by 120° in the Fig. 2 type plots (0° , 120° , 240° ; 10° , 130° , 250° , etc.). The position of the density maximum defines the group's torsional angle. The staggered conformation is defined as 0° . Plots were placed on the same relative scale by adjusting the lowest average density point for each individual group to be at 0. Effects on the methyl densities due to the inherent vibrational motion of the parent carbon atoms were corrected for by an empirical procedure. This was done by developing an empirical function relating the expected attenuation of the methyl peak density to the vibrating motion (B factor) of the parent carbon atom.

profile curve (Fig. 3) provides useful information on the rotor's characteristic minimum energy, the height and breadth of the curve indicating the rotor's degree of order. Interpretation of these plots, even at the high resolution ($1.4 \text{ \AA}$) and phasing accuracy ($R=0.17$) of this analysis, has distinct limitations. Model studies indicate that while the resolution is sufficient to assign precisely the orientation of rotor groups when a position is highly occupied, attempts to discriminate between localized vibrational or static disorder effects are not warranted¹⁵.

Rotationally averaged plots for three representative side-chain types, alanine, threonine and valine, are shown in Fig. 3. Taken together, these plots clearly suggest that the preferred conformations in the large majority of cases is closely staggered. There are, however, a few exceptions to this general observation. The clearest example of this is the CG1 rotor of Val 15 ($\triangle$, Fig. 3c) which is displaced by about 25° from the staggered orientation. This deviation is likely to represent a real effect and not an artefact because the symmetrical nature of the density peaks, each of which is displaced by about 20 – 30° , is very systematic ($\square$, Fig. 2). We believe this occurs because the CG1 methyl hydrogens make van der Waals' contact with a portion of a symmetry-related molecule in the crystal packing lattice. On the other hand, methyl groups at all other similar intermolecular contact regions do have staggered conformations, so no single generalization applies to these types of interactions.

A significant finding of this study is that a majority of the rotor density plots have distinctive 3-fold symmetry. We suggest that well-defined energy minima for methyl groups exist in protein molecules. (Similar findings were obtained from a trypsin analysis¹¹, but because the study was done at lower resolution ($1.8 \text{ \AA}$) the interpretation of the density plots was less definitive.) The staggered conformation of the methyl hydrogens is consistent with the staggered conformations seen around side-chain carbons in high-resolution protein structures determined by X rays. This is a general observation described in many studies. Experimental observations from NMR (showing rapid rotatory motions) together with the neutron diffraction results described here (indicating highly preferred orientations) show that methyl rotations are 'quantized' in 120° steps about a position of highest stability. The lifetime in any particular orientation is short in absolute terms, but it is much longer than times spent in the rotational transition from one 3-fold energy minimum to the next.

The experimental observations presented here address an important question about protein packing. Examination of space-filling models of proteins shows that the groups contained in the interior core are usually extremely close-packed with regard to neighbouring atoms and seemingly have little room to manoeuvre. It might be expected, therefore, that a number of the methyl groups on aliphatic side chains could be forced out of their stable rotatory conformations by packing considerations. Both the crambin and trypsin¹⁶ analyses, however, show that such situations are rare. This implies that the inherent 'breathing' of the protein and its small but numerous packing defects^{17,18} allow the methyl groups to adopt the low energy conformation defined by their intrinsic barrier to rotation.

We thank Dr M. Teeter for assistance in the crambin analysis, J. Shpungin and Dr B. P. Schoenborn for technical assistance, and Drs A. Kossiakoff, R. Levy and G. Rose for advice during the preparation of the manuscript. A.A.K. thanks Genentech, Inc. for support during part of the work. Research at Brookhaven National Laboratory under the auspices of the US Department of Energy was supported by NIH grant GM29616.

Received 13 April; accepted 26 July 1984.

- Gurd, F. N. & Rothgeb, T. M. *Adv. Protein Chem.* **33**, 73–165 (1979).
- Karplus, M. & McCammon, J. A. *A. Rev. Biochem.* **53**, 263–300 (1983).
- Wagner, G., DeMarco, A. & Wüthrich, K. *Biophys. Struct. Mech.* **2**, 139–158 (1976).
- Wüthrich, K. in *Nuclear Magnetic Resonance in Biological Research—Peptides and Proteins* (North-Holland, Amsterdam, 1976).
- Campbell, I. D., Dobson, C. M. & Williams, R. J. P. *Proc. R. Soc. B189*, 503–509 (1975).
- Oldfield, E., Norton, R. S. & Allerhand, A. *J. Biol. Chem.* **250**, 6368–6380 (1975).
- Wittebort, R. J., Rothgeb, T. M., Szabo, A. & Gurd, F. N. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1059–1063 (1979).
- McCammon, J. A. & Karplus, M. *Biopolymers* **19**, 1375–1405 (1980).
- McCammon, J. A. & Karplus, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3585–3589 (1979).
- Lipari, G., Szabo, A. & Levy, R. M. *Nature* **300**, 197–198 (1982).
- Kossiakoff, A. A. & Spencer, S. A. *Biochemistry* **20**, 6462–6474 (1981).
- Teeter, M. M., Mazer, J. A. & L'Italien, J. J. *Biochemistry* **20**, 5437–5443 (1981).
- Hendrickson, W. A. & Teeter, M. M. *Nature* **290**, 107–113 (1981).
- Teeter, M. M. & Kossiakoff, A. A. in *Neutrons in Biology, Basic Life Sciences* **27**, 335–348 (Plenum, New York, 1984).
- Kossiakoff, A. A. *Meth. Enzym.* (in the press).
- Kossiakoff, A. A. *A. Rev. Biophys. Bioengng* **12**, 159–183 (1983).
- Lumry, R. & Rosenberg, A. *Colloques natn. Cent. natn. Rech. scient.* **246**, 53–63 (1975).
- Richards, F. M. *Carlsberg Res. Commun.* **44**, 47–63 (1979).

'Z-RNA'—a left-handed RNA double helix

Kathleen Hall, Phillip Cruz, Ignacio Tinoco Jr,
Thomas M. Jovin* & Johan H. van de Sande*†

Chemistry Department and Laboratory of Chemical Biodynamics,
University of California, Berkeley, California 94720, USA

* Abteilung Molekulare Biologie, Max Planck Institut für
Biophysikalische Chemie, D-3400 Göttingen, FDR

In contrast to double-stranded DNA, there has so far been little evidence that double-stranded RNA can undergo major conformational transitions. We have investigated the conformation in different conditions of the double-stranded RNA molecule poly(G-C)·poly(G-C), by NMR, circular dichroism and absorbance spectroscopy. We report here evidence obtained by these different spectroscopic techniques that poly(G-C)·poly(G-C) undergoes a transition from the A-form to a left-handed Z-form in conditions of high ionic strength and at temperatures above 35 °C. This conformational transition may be of relevance to the biological situations in which double-stranded RNA occurs, such as in ribosomes and in some viruses.

Figure 1 shows the phosphorus NMR spectra of three forms of poly[r(G-C)]·poly[r(G-C)]. There are two forms of RNA that are identified by ³¹P-NMR in the lower salt concentrations, one in 10 mM HEPES where the resonances are separated by 0.40 p.p.m. (Fig. 1a) and the other in 3 M NaClO₄ where the separation is 0.50 p.p.m. (Fig. 1b). In contrast, the spectrum of poly[r(G-C)]·poly[r(G-C)] in 6 M NaClO₄ at 45 °C shows a much larger separation of 1.33 p.p.m. (Fig. 1c). This pattern is similar to that seen by Patel *et al.*¹ for poly[d(G-m⁵C)] and poly[d(G-C)] double strands.

The large separation of the phosphorus resonances is characteristic of Z-form DNA^{1,2}. The two well-separated ³¹P resonances have been assigned³ to the CpG and GpC residues, which in the crystallographic dinucleotide repeat unit have greatly different phosphodiester torsion angles⁴. This produces a zig-zag form with two distinct chemical and magnetic environments for the phosphorus. Our data suggest that the poly[r(G-C)]·poly[r(G-C)] has adopted a Z-form; this conformation will be referred to as Z-RNA.

We have used transient proton nuclear Overhauser effect (NOE) measurements to investigate the conformation about the glycosidic bond. NOE arises from through-space dipolar interactions between spins, and can be observed as a change in intensity of one peak upon irradiating another. These spins must be spatially proximal, as the magnitude of the NOE falls off as 1/*r*⁶, where *r* is the distance between the spins. In the Z-DNA structure as determined by X-ray crystallography⁴, the guanosine residues adopt a *syn* conformation about the glycosidic bond, and in this form, the guanosine H-8 (GH8) and guanosine H-1' (GH1') protons are only 2.2 Å apart. The guanosine residues adopt an *anti* conformation in A- and B-form, however, and the GH8 to GH1' distance increases to greater than 3.5 Å. Patel *et al.*¹ showed that in low salt, the only NOE from an aromatic proton in poly[d(G-C)]·poly[d(G-C)] was from the cytosine H-6 (CH6) to its neighbouring cytosine H-5 (CH5); these protons are separated by only 2.4 Å. When the polymer was placed in 4 M salt, in addition to the CH6 to CH5 NOE, an NOE was observed between the GH8 and GH1', indicating a transformation to *syn*.

Figure 2a shows the proton NMR spectrum of poly[r(G-C)]·poly[r(G-C)] in 6 M NaClO₄, 10 mM phosphate, pD = 7 at 45 °C. Only resonances due to the GH8 and CH6 protons appear in the region between 7 and 8 p.p.m., and the peak at 7.82 p.p.m. has been assigned to the GH8 using deuterium exchange at high

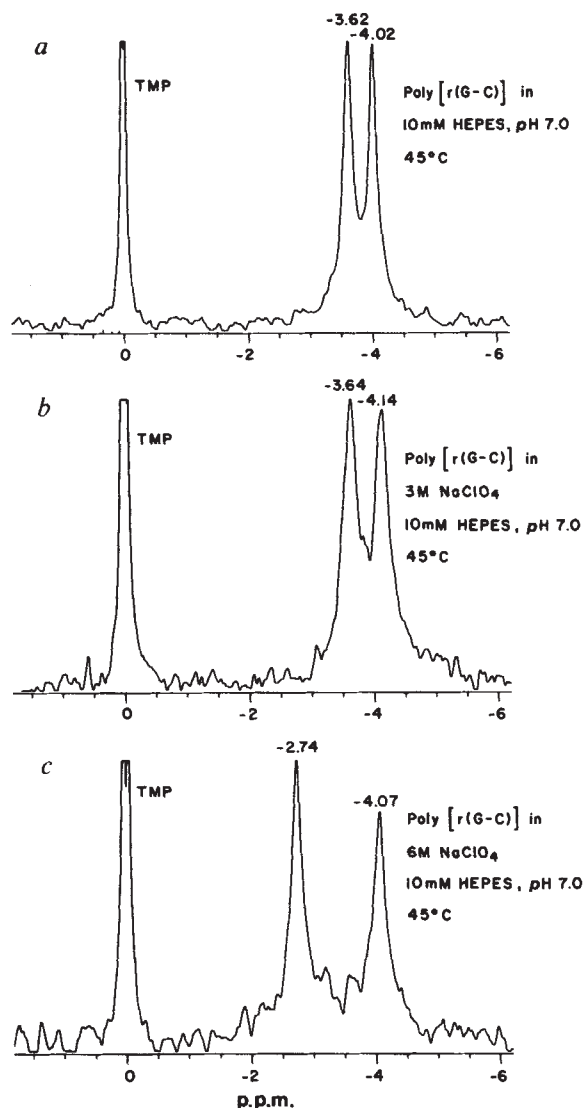


Fig. 1 Phosphorus NMR spectra of poly[r(G-C)]·poly[r(G-C)] in three ionic strengths at 45 °C. a, Poly[r(G-C)]·poly[r(G-C)] with no added NaClO₄. b, Poly[r(G-C)]·poly[r(G-C)] in 3 M NaClO₄. c, Poly[r(G-C)]·poly[r(G-C)] in 6 M NaClO₄.

Methods: The polymer was prepared using *Escherichia coli* RNA polymerase on a poly[d(I-C)]·poly[d(I-C)] template¹². Polymer size was determined by polyacrylamide gel electrophoresis in denaturing conditions. Size markers of simian virus 40 *Hinf* I digests were used for comparison. The size distribution ranged from 40 to 800 nucleotides, most being 150–500 nucleotides. The DNA template was removed by DNase I digestion; removal was complete as assessed by optical melting and CD data. Spectra were obtained at 81.75 MHz on the UCB-200 spectrometer (200 MHz proton), using broad-band proton decoupling. Each sample contained approximately 30 mM (in nucleotides) of the polymer, 10 mM HEPES, pH 7.0, 0.1 mM EDTA; chemical shifts are referenced to the internal standard, trimethyl phosphate (TMP).

temperature. The GH8 shows a large NOE to a resonance at 5.7 p.p.m., which is assigned to the GH1' (Fig. 2b). The presence of this NOE shows that the guanosine residues adopt the *syn* conformation. These NOE measurements, together with the phosphorus NMR data, provide the basis for designating the high salt form of poly[r(G-C)]·poly[r(G-C)] as Z-RNA. The CH6 proton also shows a large NOE (Fig. 2c) to a peak at 5.17 p.p.m., which has been assigned to the neighbouring CH5.

Figure 3a shows the proton NMR spectrum of poly[r(G-C)]·poly[r(G-C)] in 3 M NaClO₄, 10 mM NaPO₄, pD = 7, at 45 °C. Deuterium exchange at high temperature was used to assign the resonance at 7.51 p.p.m. to the GH8 proton. In the

† Present address: Department of Medical Biochemistry, University of Calgary, Calgary, Alberta, Canada T2N 4N1.

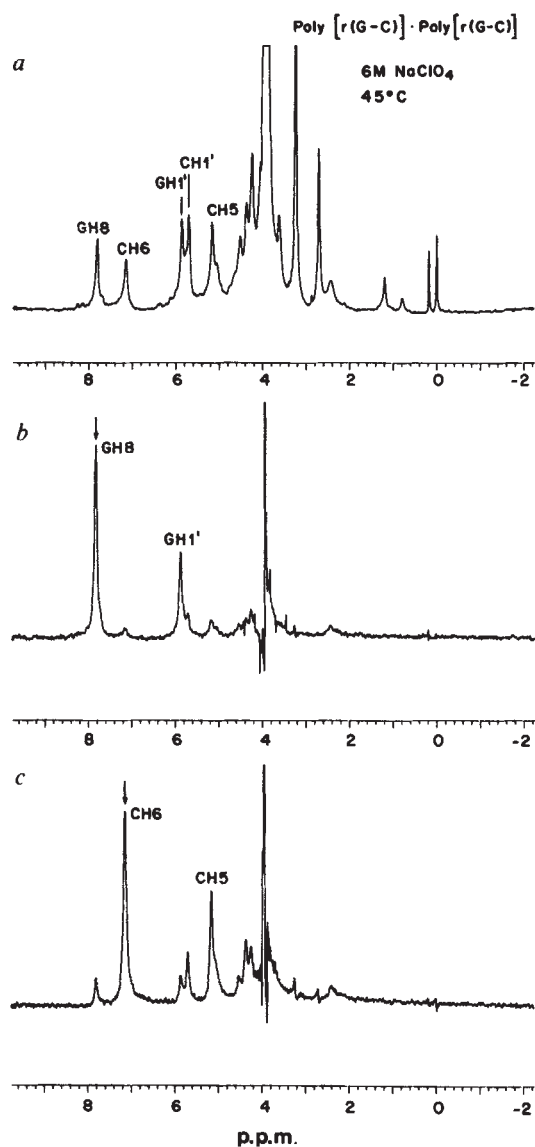


Fig. 2 Proton NMR spectra at 500 MHz of poly[r(G-C)]·poly[r(G-C)] (~30 mM in nucleotides) at 45 °C in 6 M NaClO₄, 10 mM phosphate buffer, pH 7.0, and ~0.1 mM EDTA. Chemical shifts are referenced to the internal standard trimethylsilyl propionate (TSP). Spectra were obtained on the JEOL GX-500 instrument at the Stanford University NMR Center. *a*, Spectrum in D₂O. The large narrow resonances below 5 p.p.m. correspond to residual solvent protons, EDTA and reference TSP. *b*, Transient NOE difference spectrum obtained by irradiating the GH8 proton (indicated by an arrow) with a 50-ms saturation pulse before obtaining the spectrum. The difference spectrum represents the spectrum with an on-resonance presaturation pulse subtracted from the spectrum with the presaturation pulse 10 kHz off resonance. A large NOE to the GH1' is observed. The reverse experiment (not shown) results in NOE to the GH8 proton on irradiation of the GH1'. *c*, Transient NOE difference spectrum obtained by irradiating the CH6 proton (indicated by an arrow) as described above. A large NOE can be seen to the CH5 proton. Irradiation of the CH5 proton results in a large NOE to the CH6 (not shown).

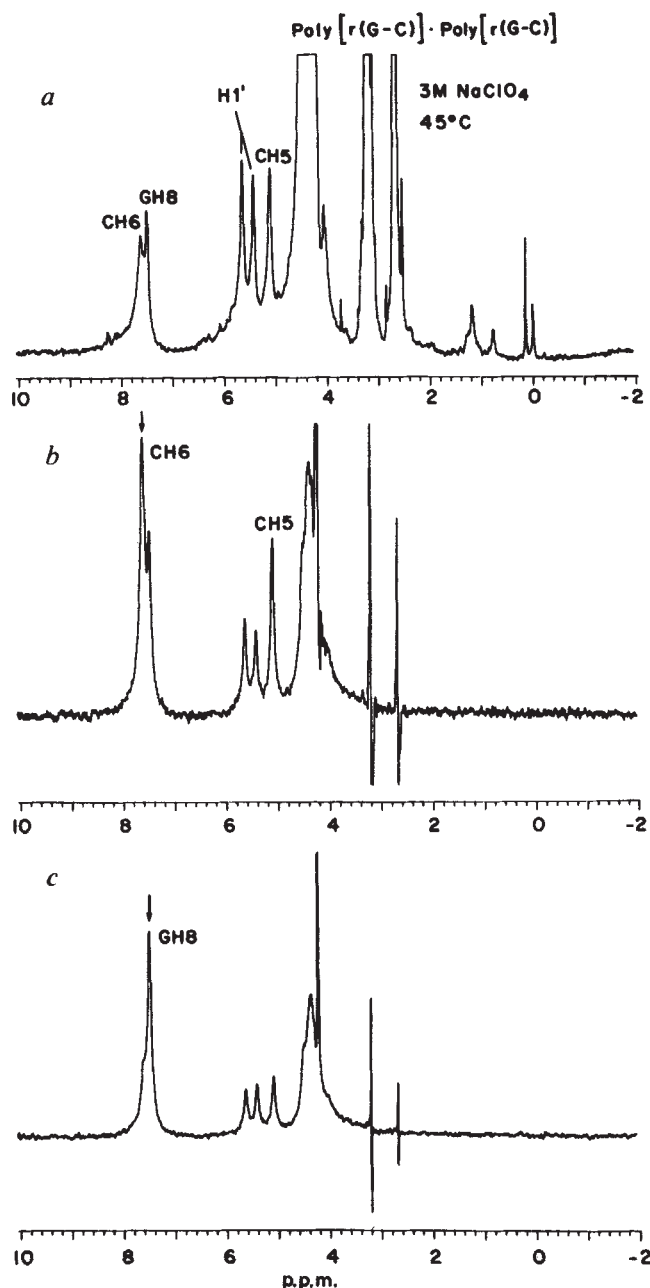


Fig. 3 Proton NMR spectra of poly[r(G-C)]·poly[r(G-C)] in 3 M NaClO₄. All other conditions are as described in the legend to Fig. 2. *a*, Spectrum at 500 MHz in D₂O. *b*, Transient NOE difference spectrum obtained by irradiating the CH6 proton (indicated by an arrow) as explained in Fig. 2 legend. A large NOE to the CH5 is observed. *c*, Transient NOE difference spectrum obtained by irradiating the GH8 proton (indicated by an arrow). Only small NOEs due to spin diffusion are observed.

CH6 peak at 7.63 p.p.m. results in a large NOE to a peak at 5.11 p.p.m.; this peak has been assigned to CH5. When the GH8 peak at 7.51 p.p.m. was irradiated, only several small NOEs were observed, indicating that guanosine is in the *anti* conformation. It is seen from the NOE spectra that there is a significant amount of spin diffusion among the various protons. This was also seen with poly[d(G-C)] (ref. 1) and can to some extent be diminished by shortening the duration of the saturation pulse. It was not entirely eliminated in these experiments, however, and appears as small NOEs in all the spectra.

The absorption and the circular dichroism (CD) spectra of two forms of RNA (Fig. 4) corroborate the NMR conclusions. The Z-RNA absorption spectrum (6 M NaClO₄, 45 °C) has a

A-form, exchange is extremely slow; this is in contrast to the Z-form, which exchanges more rapidly. Therefore, to promote deuterium exchange, the poly[r(G-C)]·poly[r(G-C)] was heated to 80 °C for several hours in 1 mM Na phosphate, 0.1 mM EDTA.

Figure 3*b, c* shows the NOE difference spectra of poly[r(G-C)]·poly[r(G-C)] in 3 M NaClO₄ that result upon irradiation of the peaks at 7.63 and 7.51 p.p.m., respectively. Saturating the

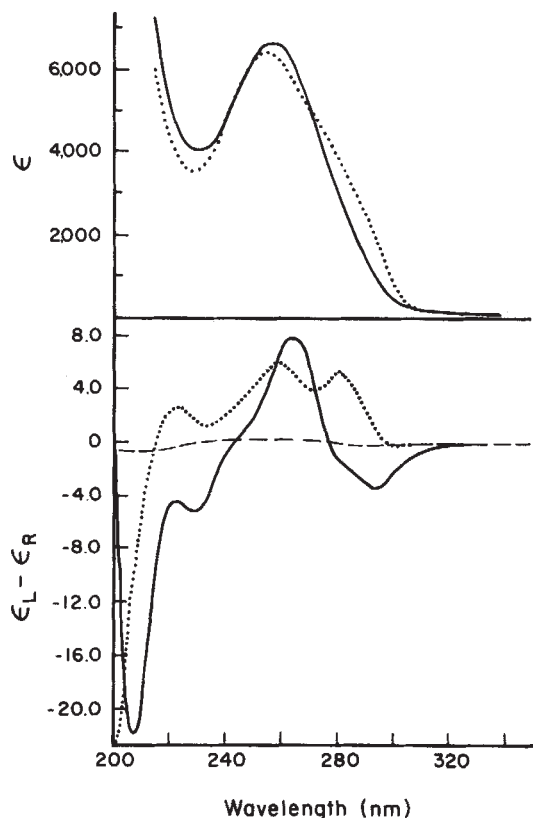


Fig. 4 Absorption (top) and CD (bottom) spectra of two forms of poly[r(G-C)]·poly[r(G-C)]. Absorption spectra were measured with a Gilford model 250 spectrophotometer; CD spectra were measured with a JASCO J500C spectropolarimeter. Both instruments have thermoelectric cell blocks. The absorbance and CD were measured in 10 mM phosphate, pH=7.0, 6 M NaClO₄ at 22 °C (solid line) and 45 °C (dotted line). An extinction coefficient of 6,560 at 260 nm was used for the poly[r(G-C)] at 22 °C (ref. 6). The CD spectrum at 22 °C is similar, but not identical to the spectrum seen in buffer alone. It is identical to that for the polymer in 3 M NaClO₄. At 45 °C, 6 M NaClO₄ the CD spectrum is identical to that seen in conditions of lower perchlorate and higher temperatures, and is similar to that in 20% (v/v) ethanol in 4.8 M NaClO₄.

shoulder at 290 nm; this is very similar to the spectrum of Z-DNA (ref. 5). Two types of CD spectra are seen for poly[r(G-C)]·poly[r(G-C)]. The first is represented by the spectrum at 6 M NaClO₄, 22 °C. It is identical to the CD spectrum in 3 M NaClO₄ and is similar to a CD spectrum measured in 10 mM sodium phosphate (not shown) which matches the CD spectrum of Gray *et al.*⁶ in 1 mM Na phosphate, 0.1 mM EDTA, pH 7.8. These CD spectra are all presumed to be from an A-type conformation of the poly[r(G-C)] (ref. 7). Because the CD spectrum of the polymer in 6 M NaClO₄, 22 °C and 3 M NaClO₄ are identical, we used the latter solvent at 45 °C for NMR experiments; at 22 °C the NMR linewidths are too broad for analysis. The CD spectra in dilute phosphate differ from those in NaClO₄ by a small change in the band at 285 nm. This band is particularly sensitive to solvent, and diminishes at high ionic strength. This CD band may be correlated with the small differences in phosphorus resonances between the RNA in 10 mM HEPES and in 3 M NaClO₄.

The CD spectrum in 6 M NaClO₄, 45 °C is very different from the others, and is identified as that of Z-RNA. The CD remains positive in the region above 215 nm. The negative bands above 280 nm and below 240 nm in A-RNA change sign in Z-RNA; the positive band near 260 nm remains positive.

The transition temperature for the A- to Z-RNA transition is a function of the solvent conditions. In 6 M NaClO₄, the transition temperature is ~35 °C; in 4.8 M, it is ~45 °C; in 4 M, ~80 °C; in 3 M, it is above 80 °C and not measurable in our

instrument. The addition of ethanol to the perchlorate solution decreases the transition temperature, so that in 4.8 M NaClO₄/20% ethanol, the RNA is in Z-form at room temperature. This temperature dependence of the transition to Z-RNA form is analogous to that of poly[d(G-C)]·poly[d(G-C)]. For that polynucleotide, the Z-form is favoured at higher temperatures in LiCl or ethanol/water solutions^{1,8,9}. However, we have found that solvents which will cause a B to Z transition in poly[d(G-C)], such as 4 M NaCl or 1 M MgCl₂, do not cause an A to Z transition in poly[r(G-C)]. This is consistent with the work of Westerink *et al.*¹⁰ and Uesugi *et al.*⁷ who found no Z transition in these salt solutions for r(G-C) oligonucleotides.

The kinetics of the transition also vary with the solvent and the temperature. In 6 M NaClO₄, a temperature shift from 22 °C to 35 °C results in an A- to Z-RNA transition which takes ~1 h. The reverse transition occurs over 5 h. A temperature shift from 22 °C to 45 °C (or addition of 20% (v/v) ethanol and concomitant dilution to 4.8 M NaClO₄), however, produces the Z-RNA form in ~10 min, although again the reverse transition on lowering the temperature from 45 °C to 22 °C takes several hours.

This is the first major conformational change reported for a double-stranded RNA; it shows that RNA has conformational lability, contrary to current belief. It is consistent with the finding of Uesugi *et al.*¹¹ that r(C-m⁸G-C-m⁸G) exists in a Z-like structure. The transition to Z-form RNA requires more extreme conditions than the transition to Z-form DNA. However, in analogy to Z-DNA, other more physiological environments may be found to favour Z-RNA, and other RNA sequences may also undergo the transition. In any case one should now consider conformations other than A-form when considering double-stranded regions of RNA in ribosomes, viruses and other RNA-protein complexes.

We thank Professor Michael Chamberlin and his laboratory for providing us with RNA polymerase as well as advice on the synthesis and purification of poly[r(G-C)]·poly[r(G-C)]. We also thank Dr Dave Wilbur for assistance on the GX-500 spectrometer at Stanford. The proton NMR experiments were performed at Stanford University NMR Center, which is supported by NSF grant GP23633 and NIH grant RR00711. The work at Berkeley was supported in part by NIH grant GM10840 and the US Department of Energy, Office of Energy Research, under contract 03-82ER60090.000.

Received 4 June; accepted 9 August 1984.

1. Patel, D. J., Kozlowski, S. A., Nordheim, A. & Rich, A. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1413-1417 (1982).
2. Patel, D. J., Canuel, L. L. & Pohl, F. M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2508-2511 (1979).
3. Jovin, T. M. *et al. Cold Spring Harb. Symp. quant. Biol.* **47**, 143-154 (1983).
4. Wang, A. H.-J. *et al. Science* **211**, 171-176 (1981).
5. Pohl, F. M. & Jovin, T. M. *J. molec. Biol.* **67**, 375-396 (1972).
6. Gray, D. M., Liu, J.-J., Ratliff, R. L. & Allen, F. S. *Biopolymers* **20**, 1337-1382 (1981).
7. Uesugi, S. *et al. J. biol. Chem.* **259**, 1390-1393 (1984).
8. van de Sande, J. H. & Jovin, T. M. *EMBO J.* **1**, 115-120 (1982).
9. Hall, K. B. & Maestre, M. F. *Biopolymers* (in the press).
10. Westerink, H. P., van der Marck, G. A., van Boom, J. H. & Haasnoot, C. A. G. *Nucleic Acids Res.* **12**, 4323-4338 (1984).
11. Uesugi, S. *et al. J. Am. Chem. Soc.* **106**, 3675-3676 (1984).
12. Hall, K. B., Cruz, P. & Chamberlin, M. F. *Archs Biochem. Biophys.* (in the press).

Erratum

Dynamics of meltwater discharge from Northern Hemisphere ice sheets during the last deglaciation

R. H. Fillon & D. F. Williams
Nature **310**, 674-677 (1984)

IN the above letter, the last four lines of the penultimate paragraph on p. 676 ('anomaly areas... $\delta^{18}\text{O}_{\text{sw}}$ anomaly') are a repetition from a previous paragraph. The correct four lines read:

amount of meltwater discharged into the Canadian Current before 13 kyr BP according to our calculations. This excess may reflect an early marine-drawdown phase of North American deglaciation^{4,23}.

Preformation to punctuation

Colin Patterson

Evolution: The History of an Idea. By Peter J. Bowler.
University of California Press: 1984. Pp. 412.
Hbk \$34.45, £23.95; pbk \$12.60, £8.75.

I CAME to this excellent book with vague misgivings, remembering the tedium of Henry Fairfield Osborn's *From the Greeks to Darwin*, or Edward Clodd's *Pioneers of Evolution*. No need to worry, for neither is cited in Peter Bowler's massive 54-page bibliography, and history of science is long past relentless progress towards today's truth, as chronicled in those older works. But Bowler does say that

A few years ago, it would have been difficult to write a history of evolutionism except from the perspective that the modern form of Darwinism represents a triumphant climax to the process.

To my real surprise, I learned from Bowler's bibliography that his book has no real competitor — there is no modern, general history of evolutionism. The closest competitors, I suppose, are D.R. Oldroyd's *Darwinian Impacts* (Open University Press, 1980), and Ernst Mayr's magnum opus, *The Growth of Biological Thought* (Harvard University Press, 1982). In his opening chapter Bowler notes, and denies, the distinction between "internalist" and "externalist" historians of science, the first studying the "objective process of scientific discovery", the second looking at "how society reacted". He argues that the distinction disappears once we realize that scientific knowledge is not purely objective; perhaps, but one would have to rate Oldroyd's book as more externalist than Bowler's, and Mayr's as more internalist. Both are more difficult going. Bowler aims at readers with no background in either biology or history, especially students in history of science, but I believe that most biologists and many other scientists could read him with pleasure and profit.

The book is finely and carefully balanced, as exemplified by the arrival of *On the Origin of Species* exactly midway through the text, so that half the book is pre-1859, and half after that date. Bowler has particular strengths in both halves, for he has published in history of biology from seventeenth-century theories of reproduction to twentieth-century genetics, notably in books on nineteenth-century palaeontology (*Fossils and Progress*; Science History Publications, 1976) and on "anti-selectionism" around 1900 (*The Eclipse of Darwinism*; Johns Hopkins University Press, 1983). Throughout the book, I was struck by the way Bowler compresses complicated ideas into a page or two, and still finds space for illuminating comment: someone who can do this even for so thoroughly studied a

figure as Darwin (for example on pp. 142–146) is doing excellent work.

There are two technical blemishes, the first trifling, of the kind it pleases pedants to discover — the portrait on the dust-jacket, said to be of J.B.S. Haldane, is his father J.S. The second blemish, more intrusive, is Bowler's usage of the word "hierarchy", a vital and recurring concept in the history of evolution. He uses "hierarchy" to mean linear sequence, as in the "chain of being" or a chain of command, where the general's orders pass down the military hierarchy to the private. But in biology today, "hierarchy" entails the inclusion relation, so that the higher includes the lower rather than outranks it: this usage is definitive in taxonomy, but also general in ontogenetic theory and other fields invoking theories or processes of branching. Bowler's idiosyncratic or old-fashioned usage leads to apparently nonsensical statements such as "Linnaeus' system [of classification]...does not imply a hierarchical ranking" or "Cuvier's reluctance to accept hierarchical classification"—here the notion of branching, or inclusion, is seen as the antonym of hierarchy.

The book ends with a chapter on "The

Modern Debates", picking out four or five areas of controversy. Surprisingly neutralism, and the associated "molecular clock", get no mention here. Indeed, neutralism receives very short shrift in the book, three passing mentions, one of which is mistaken:

If the survival of the fittest were necessarily true, it would be impossible for anyone to imagine that it was false—yet there have been theories of evolution based on random factors that deny the superior reproductive power of the fit.

The mistake here is to equate natural selection with change (directional selection), whereas neutralists emphasize the dominance of stabilizing selection, which resists change. Neutral theory seems to me to deserve at least as much space as punctuated equilibrium, which gets a very fair hearing, and modern Creationism, which gets 11 pages. But one point in those pages deserves thought:

It may sound paradoxical, but the only way creationists can be forced to expose their own extraordinary theory to close analysis is if evolutionists concede that they cannot prove their own case. It is only by admitting the lack of certainty inherent in any theory that we can shift the argument to a new level.

Bowler occasionally slips from that noble level, as with "The true story of the horse's evolution has turned out to be rather more complicated" and similar comments implying access to the truth on human evolution. But all in all, I salute what I believe is a superb job. □

Colin Patterson is Senior Principal Scientific Officer in the Palaeontology Department, British Museum (Natural History), London.

Records of the early metazoans

Steven M. Stanley

The Dawn of Animal Life: A Biohistorical Study.

By Martin F. Glaessner.
Cambridge University Press: 1984.
Pp. 244. £25, \$49.50.

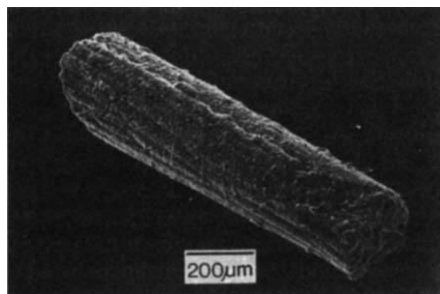
FOR nearly a century after Darwin published *On the Origin of Species*, the nature of Precambrian life was debated fruitlessly in a climate of empirical uncertainty. Early fossil finds were continually announced and continually debunked, so that in any given year some claims were alive but clearly vulnerable to rejection.

During the 1950s and 1960s, the discovery of four types of Precambrian and earliest Cambrian fossils altered the picture dramatically, adding greatly to the temporal range of palaeontological inquiry. The four new fossil groups were single cells preserved in fine-grained

sedimentary rocks; imprints of soft-bodied animals; tiny skeletal fossils slightly older than the first Cambrian trilobites; and trace fossils (tracks, trails and burrows). Of the four, single cells have often been seen as the most exciting finds, partly because their record extends some 2×10^9 years further back in time than the record of multicellular life. In many ways, however, the study of Precambrian cells has been disappointing: increase in cell size and surface complexity about 1.4×10^9 years ago point to the primary expansion of eukaryotes at this time, but the fossil record reveals little about the biological nature of earlier prokaryotic cells. The early metazoan record has been far more revealing, and we must welcome Martin Glaessner's publication of a book that focuses on this record, pulling together our knowledge, while briefly reviewing Precambrian microfossils. Glaessner, a pioneer in the study of Precambrian metazoans, has kept abreast of new developments, including many in the Soviet Union that have gone largely unnoticed elsewhere.

As might be expected, Glaessner devotes about half of his book to the spectacular

Ediacara soft-bodied fauna of Australia, which he first elucidated in 1958 and 1959, and to similar but less rich faunas that have since been discovered in many other parts of the world. The treatment of these faunas is the most comprehensive yet and is perhaps weakened only by the paucity of photographs — written descriptions of important morphological and preservational features leave the reader no recourse but to seek illustrative material



Courtesy M.D. Brasier

Part of the picture of early life — *Coleoloides typicalis*, a calcareous tubular fossil from the earliest Cambrian Hyolithes Limestone of Nuneaton, UK.

among the other works cited in the book. In Glaessner's interpretation, the Ediacara fauna is dominated by medusoid and sea-pen-like coelenterates as well as annelids and arthropods. Unfortunately, the book was published too soon for the author to respond to Professor A. Seilacher of Tübingen, who is currently challenging these taxonomic assignments. Perhaps the need for a response will prompt Professor Glaessner to provide another review with plentiful photographs of high quality. Glaessner's global reviews of Late Precambrian trace fossils and Tommotian (earliest Cambrian) shelly fossils are also comprehensive and authoritative, though unfortunately they are even less well illustrated.

A nagging question that remains in Precambrian palaeontology is: when did the metazoans actually originate? Trace and body fossils reveal the presence of metazoans in considerable diversity after the end of the Varangian glaciation, about 6.5×10^8 years ago, but what about an earlier history? Glaessner concludes that the metazoan record extends to about 1×10^9 years before the present, but the evidence he cites is meagre and equivocal. If some of the alleged trace fossils are real, the suppression of metazoan diversification for several hundred million years remains one of the greatest puzzles in palaeontology.

Although Glaessner has been unable to shed much light on this problem, his book does provide a unique, up-to-date evaluation of Late Precambrian life. It should expand the knowledge of all palaeontologists, save a few specialists, and serve as a basis for graduate seminars on the early diversification of metazoan life. □

Steven M. Stanley is Professor of Paleobiology at The Johns Hopkins University.

Gauge theories with a difference

Graham G. Ross

Quarks and Leptons: An Introductory Course in Modern Particle Physics.

By Francis Halzen and Alan D. Martin.
Wiley: 1984. Pp. 396. Hbk £35.45, \$31.95; pbk £12.50, \$13.75.

Gauge Theories of the Strong, Weak, and Electromagnetic Interactions. Frontiers in Physics.

By Chris Quigg.
Benjamin-Cummings/Addison Wesley: 1984. Pp. 334. \$42.95, £33.60.

Gauge Theory of Elementary Particle Physics.

By Ta-Pei Cheng and Ling-Fong Li.
Clarendon: 1984. Pp. 450. Hbk £35, \$39.95; pbk £15.

Gauge Theories of Strong and Electroweak Interactions.

By P. Becher, M. Böhm and H. Joos.
Wiley: 1984. Pp. 306. £28, \$54.95.

It is indicative of the state of high energy physics that four textbooks should appear together covering broadly the same ground. In recent years gauge field theories have emerged for the strong and the weak interactions, complementing quantum electrodynamics, the remarkably successful theory of the electromagnetic interaction. A large amount of experimental data has now accumulated consistent with the predictions of these theories, culminating with the detection of the gluons (the strong interaction partners of the photon) and of the $W^\pm$ and Z intermediate vector bosons (the weak interaction partners of the photon).

After the success of these theories there has been something of a lull in theoretical activity, leaving time for reflection and organization. The four books reviewed here are products of this period and all introduce the subject of local gauge field theories and their application in particle physics. However each book has a character of its own and addresses a different level of audience.

Pitched at an introductory level is the book by Halzen and Martin. It acquaints the reader with the nature of the fundamental interactions and the fundamental building blocks of matter, quarks and leptons, and is at its best in the explicit treatment of the use of Feynman rules to calculate physical processes at tree level. Many worked examples are given (with all the steps) and the text is supplemented by a large number of instructive problems; together, these features should enable readers with no prior knowledge in the subject to calculate simple processes for themselves.

The treatment of fermions and the early introduction of a chiral basis for Dirac matrices is clear and thorough, but

somewhat less successful is the treatment of the group theory which falls short of being really useful. Beyond the tree level the book is, possibly necessarily, less explicit. While I was disappointed that the introduction of local gauge theories and the Lagrangian formalism was left to the penultimate chapter — too late for a thorough treatment — I liked this book in general. Above all it should prove to be a useful part of an introductory course in particle physics.

In many ways, Chris Quigg starts where Halzen and Martin leave off. After a nice introduction to quarks, leptons and gauge bosons, local gauge field theories are introduced via the Lagrangian formalism. In the main, the reader is expected to be able to calculate Born graphs, leaving space for a fuller discussion of quantum chromodynamics (QCD — the gauge theory of the strong interactions) and of spontaneously broken gauge theories used in describing the weak interactions.

There is a good discussion of radiative corrections in QCD, well motivated by explicit calculations and leading to a Feynman graph analysis of the leading log summation, but it stops short of the renormalization group analysis needed to justify and simplify this summation. The general emphasis is on explicit physical applications, the book containing good coverage of deep inelastic scattering, $\gamma\gamma$ scattering and e^+e^- annihilation. The treatment of the electroweak theory is easy to follow but is only at tree level; no attempt is made to present a loop calculation in a spontaneously broken gauge theory. The book ends with a brief but helpful discussion of the extension to grand unified theories, in particular to SU(5). Throughout problems are set, carefully chosen to augment and illustrate the textual discussion, while the extensive reference list will help readers delve further into particular topics. Overall this is a good introductory book, with plenty of physical insights, and will repay study.

By far the longest book of the four is that by Cheng and Li. It is also the most ambitious in terms of the subject matter covered, with emphasis more on the theoretical aspects. Part One covers the formal aspects of field theory, group theory and chiral symmetry in strong interactions, as well as the construction of local gauge field theories, their Feynman rules and the calculation of radiative corrections. All this should be useful to theoretical students with some prior knowledge for it neatly summarizes a vast subject. Because of its wide scope, however, there is necessarily little time for elaboration and I think this part would be heavy going for a newcomer to the field. Renormalization and the renormalization group are discussed in detail and this aspect of the book would complement the more physical analysis in terms of Feynman graphs presented, for example, by Quigg.

In Part Two, Cheng and Li move on to

QCD and the electroweak theory. I particularly appreciated inclusion of an explicit treatment of loop calculations in a spontaneously broken gauge theory, and also the detailed discussion of the Kobayashi-Maskawa matrix, CP violation and neutrino masses. All in all the book contains a lot of goodies, including grand unification and nonperturbative aspects of gauge theories; but, again, one probably needs some experience to appreciate them, for often an historical approach is followed which hinders the logical development of the subject. For example, a detailed discussion of the role of light-cone singularities in the Bjorken scaling region precedes the more modern and more physical momentum space Feynman graph analysis. Also there are no set problems to guide the reader. Ideally, then, Cheng and Li's book should be read after one (or both!) of the more introductory texts discussed above.

The fourth book, by Becher, Böhm and Joos, is a translation of the German edition

which first appeared in 1981. It is different in character to the others, delving deeper into the theory of the renormalization of the gauge field theory. This is more in the nature of an advanced text or review for an experienced theoretician and contains extensive discussions, without full details of intermediate steps, of perturbative and nonperturbative aspects of QCD. The account of the electroweak theory, on the other hand, is briefer but more straightforward with a nice discussion of the high energy behaviour of Yang-Mills gauge theories.

Although the four texts reviewed here have as a general theme the gauge theories of the fundamental interactions, they vary widely in their demands on the reader. To a large extent they complement one another, and together represent a substantial improvement in the literature surveying modern particle physics. □

Graham G. Ross is in the Department of Theoretical Physics at the University of Oxford.

Pastures of muscle physiology

D.R. Wilkie

Muscles, Reflexes, and Locomotion.

By Thomas A. McMahon.

Princeton University Press: 1984. Pp.331. Hbk \$50, £50; pbk \$15, £12.50.

THIS is a brilliant book. Dr McMahon's original training as an engineer is an excellent one for a physiologist; in both branches of science we are concerned with machines that must function in an optimal way. Here we are provided with a clear and accurate summary of many of the mathematical aspects of muscular function that otherwise demand a laborious appraisal of a large number of original papers. Our understanding (and diligence) are repeatedly challenged by the examples and questions to be found at the end of every chapter.

McMahon's wide and scholarly approach to his subject, and his limpid and often witty style of writing, enable him to illuminate a whole range of topics ranging in physical scale (though not necessarily in complexity!) from the mode of operation of the individual cross-bridges thought to generate mechanical power at the molecular level, to the problems inherent in designing an elephant and in building a "tuned" running track on which athletes have consistently produced improved performances.

Naturally I do not agree with every word in the book. For example, McMahon writes (p.35) "The discovery of a process which apparently absorbs heat, like the mechanism of a refrigerator, provided a great deal of entertainment in the muscle

literature for years". Alas he does not make it nearly clear enough for my liking that the "entertainment" resulted in deep confusion which lasts to this day. There are no circumstances in which a muscle acts as a refrigerator, which is a machine that moves heat from a lower to a higher temperature by the continued application of mechanical or electrical power. Nor can I see how a "unifying principle" (p.79) is provided by the non-uniformity of sarcomere lengths. The question of the instability of stretched muscle fibres, whose tension-length curve has a negative slope, has been discussed for many years yet remains controversial. In whole anatomical muscles (rather than in isolated single fibres) an important role, probably a stabilizing one, may well be played by the connective tissue between the fibres which is plentiful and unlikely to be arranged at random. Microanatomical studies are much needed; until they have been carried out, comparison between isolated muscle fibres and intact anatomical muscle demands great caution.

The biochemical aspects of muscular function are touched on only lightly (nobody should expect everything!) but McMahon is careful to differentiate between the chemical processes and the physical energy — heat and work — that results from them. A widespread failure to make this distinction is at the root of the continuing confusion about a topic introduced on page 211, that of positive and negative work. In movements of a whole animal, the function of muscles as brakes absorbing mechanical work is almost as important as their function as motors producing mechanical work. Unlike mechanical brakes, however, the muscles consume fuel in performing their braking action and even in sustaining a static load. To pursue the mechanical analogy further,

a motor car can stop indefinitely on a hill by applying the brake and stopping the engine. A muscle must rev the engine and slip the clutch. The definition of "efficiency" given in Eqn 8.3 is ambiguous concerning positive and negative signs, though they can be inferred from the left-hand lines in Fig. 8.21 which do indicate negative values of "efficiency". This is formally correct though perhaps not very helpful. I hope that this section will make the topic as clear to the uninitiated as I know it to be to the author himself.

The last chapter, "Effects of Scale", is a delight to read and may well come as a surprise to biologists who work on subjects other than muscular contraction. Experimentally, we find in animals of different size a large number of functions that depend in a highly regular way on body weight, but in hardly a single case is there a satisfactory physical or biochemical explanation for the dependence. McMahon explores various possibilities in a way that is both instructive and entertaining. The relationship between metabolic rate and body weight is the best-established of such laws and is referred to briefly on page 279. It is known to apply to living creatures over a weight range of 10^{18} , a respectable number even by astronomical standards, yet it totally lacks an adequate explanation. This seeming vacuum is a constant source of irritation to all biologists who concern themselves with such issues. Presumably because of their recent publication dates, no references are given to the fuller analyses of the problems of scale given in the beautifully-illustrated books by McMahon and John Tyler Bonner (*On Size and Life*; W.H. Freeman, 1983) and by Knut Schmidt-Nielsen (*Scaling: Why is Animal Size so Important?*; Cambridge University Press, 1984).

Students of muscle physiology at all levels will benefit from reading *Muscles, Reflexes, and Locomotion*. Not only is it a compact source of information, but a novel, comprehensive and readable book which tempts one to browse in the wider pastures opened up by Dr McMahon. □

D.R. Wilkie is Jodrell Research Professor of Physiology at University College London.

New in paperback

- *The Enchanted Ring: The Untold Story of Penicillin* by John C. Sheehan (reviewed in *Nature* 300, 126; 1982). Publisher is MIT Press, price is \$6.95, £7.25.
- *Quarks: The Stuff of Matter* by Harald Fritzsch (reviewed in *Nature* 306, 508; 1983). Publisher is Penguin, price is £4.95.
- *W.H. Hudson: A Biography* by Ruth Tomalin, first published by Faber & Faber and reviewed in *Nature* 301, 446; 1983. Publisher of the paperback is Oxford University Press, price is £3.95.
- *Man and the Natural World: Changing Attitudes in England 1500-1800* by Keith Thomas (reviewed in *Nature* 303 641; 1983). Publisher is Penguin, price is £4.95.

Job prospects in the United States

from Richard Pearson

With the United States' economic boom continuing, and the rate of job generation far outstripping that of other Western economies, what are the job prospects for life scientists and engineers?

PROJECTING future employment levels is notoriously difficult, yet for occupations requiring many years of training and experience, insights into future employment patterns and imbalances can show where significant adjustments might be needed in labour market and training policies. Any forecasts, however, can only be as good as the assumptions on which they are based, and these need to involve assessments of future levels of economic and investment activity, and technological and productivity changes.

Although scientists, engineers and technicians account for only a small fraction of the US workforce (less than 5 per cent), they have a disproportionate role in economic growth, technical innovation and national security and are a key area of concern. A key role for the governmental National Science Foundation (NSF) in Washington is to provide information on the availability, both current and projected, of such scientific and technical resources in the United States. A recent NSF study looks at the ability of the labour force to fuel the expected 45 per cent increase in defence expenditure over the period 1982 to 1987 (see Table 1). A series of scenarios was used, looking at the implications of high and low growth rates in Gross

economy, which has led to a significant shift in staffing patterns towards SET employment. Interestingly, although each scenario focuses on rather different sets of assumptions, the projected range of SET requirements is rather more restricted, never varying by more than 5 per cent from the average 1987 projected value in any of the major occupational categories (Table 1). This is in part because SET manpower is only a small proportion of all jobs in the economy but also because productivity increases at a much faster rate under the high growth scenario, with the productivity gains counterbalancing the employment impact of economic growth which in turn reduces the need for additional SET manpower.

There were nearly 750,000 university degree level science jobs in the US economy in 1982 and this figure was expected to grow by between 116,000 (3 per cent per year) and 161,000 (4.1 per cent per year) jobs over the period to 1987. Both these figures represent a slow-down in the 8.7 per cent average annual growth rate over the period 1977–1982. The key determinants of these growth patterns are shown to be primarily economy-related, with the key driving forces for increased scientific jobs being the service sectors, particularly research and development laboratories, consulting firms and computer establishments.

The engine of this growth in science jobs seems to be the computer, with computer systems analysts, an occupation first recognized officially in the 1970 census, accounting for about 300,000 people in 1987, regardless of the scenario followed, with half of them employed in business services. The growth in these jobs will average about 6 per cent per year, the net growth being 70,000–85,000 jobs over the five years. Computer specialists are expected to dominate science jobs in 1987, increasing their share over the period from 34 to 55 per cent, with psychologists, chemists and mathematicians being the other largest groups.

Of the physical sciences, physics has the fastest growth at 1.4 to 3 per cent per year with growth for chemists being rather lower, at 1.5–2.5 per cent, a figure similar to that for geologists. Mathematical science is projected to have rather faster growth at 2.2–3.5 per cent per year but the total job gain will only be 5,000–10,000 jobs overall. In the life sciences, agricultural jobs barely keep pace with overall employment growth, and while the number of jobs for biological scientists will grow rather faster, the total job gain will be only

about 5,000–10,000 extra jobs.

By way of contrast, engineering shows an overall growth of between 2.6 and 4.5 per cent with aeronautical/astronautical employment rising by a massive 5.9–11.1 per cent per year, a gain of 22,000–45,000 jobs. Electronics is the other major growth occupation, and although it shows rather less spectacular growth of between 3.9 and 5.1 per cent, the expected gains of between 69,000 and 94,000 jobs match the absolute growth in jobs for computer systems analysts.

As well as considering possible demand, NSF also makes corresponding supply projections, taking account of such factors as new entrants to the labour force, occupational mobility, immigration, and deaths and retirements. The model projects that supply will exceed demand under all the scenarios for all science occupations; for example, there could be three people seeking every job. However, significant shortages are indicated for aeronautical/astronautical engineers and electronic engineers, together with particularly large shortages in the computer speciality fields. These projections, however, assume no major forms of labour force adjustment such as increased salaries, reduced hiring standards, significant mobility from other occupations or increased immigration. One particular concern is the possibility that industry may be so attractive that the universities will be unable to attract enough talented people for research.

Labour market adjustments are expected to alleviate the likely level and impact of shortages in the above disciplines, and also mitigate the size of surpluses. But NSF's main concern is that the larger the adjustment needed, the greater the impact on and likely lowering of the quality of the workforce. This could induce productivity and quality losses in the short term, while hindering longer-term supply of skilled manpower if the academic sector cannot retain graduate students or its existing faculty. Already many specialists in information technology are leaving academic posts for industry. However, for many of those seeking a science-linked job, employment prospects are declining and may continue to deteriorate for the next few years. Occupational mobility and retraining will then become higher priorities for the individual seeking secure employment. □

Table 1 Projected US employment changes for the years 1982–87

	Employment (thousands)		
	1982	1987 (low growth)	1987 (high growth)
Computer scientists	219	287	303
Other scientists	508	556	585
Engineers	1,140	1,296	1,423
Technicians	1,468	1,649	1,760

Values for 1982 are 'actual', and those for 1987 are projected. Source: *Projected Response of the Science Engineering, and Technician Labor Market to Defense and Non-Defense Needs: 1982–87* (National Science Foundation, Washington, 1984).

Domestic Product between 2 and 4 per cent per year, and of two defence expenditure scenarios averaging 3 and 8 per cent per year growth.

Growth in demand for scientific engineering and technician (SET) manpower is expected to grow by between 2.6 and 4.1 per cent per year, a rate of growth higher than for employment generally (up 1–2.4 per cent per year). This increasing demand is driven not only by the strong performance of high technology industries, with their high concentrations of SET manpower, but more importantly by the diffusion of technology throughout the

Richard Pearson is at the Institute of Manpower Studies, Mantell Building, University of Sussex, Brighton BN1 9RF, UK.